

# WANDERING IN CITIES

*A statistical physics approach to urban theory*

RÉMI LOUF



*Jazz is not dead,  
It just smells funny.*

—

Frank Zappa, *Be-Bop Tango* (1973)





À mes parents,  
Qui ont toujours placé l'éducation avant tout.

À tous mes enseignants :  
Votre main n'était pas tendue en vain.



## ABSTRACT

---

The amount of data that is being gathered about cities is increasing in size and specificity. However, despite this wealth of information, we still have little understanding of what really drives the processes behind urbanisation. In this thesis we apply some ideas from statistical physics to the study of cities.

We first present a stochastic, out-of-equilibrium model of city growth that describes the structure of the mobility pattern of individuals. The model explains the appearance of secondary subcenters as an effect of traffic congestion. We are also able to predict the sublinear increase of the number of centers with population size, a prediction that is verified on American and Spanish data.

Within the framework of this model, we are further able to give a prediction for the scaling exponent of the total distance commuted daily, the total length of the road network, the total delay due to congestion, the quantity of CO<sub>2</sub> emitted, and the surface area with the population size of cities. Predictions that agree with data gathered for U.S. cities.

In the third part, we focus on the quantitative description of the patterns of residential segregation. We propose a unifying theoretical framework in which segregation can be empirically characterised. We propose a measure of interaction between the different categories. Building on the information about the attraction and repulsion between categories, we are able to define classes in a quantitative, unambiguous way. The framework also allows us to identify the neighbourhoods where the different classes concentrate, and characterise their properties and spatial arrangement. Finally, we revisit the traditional dichotomy between poor city centers and rich suburbs; we provide a measure that is adapted to anisotropic, polycentric cities.

In the fourth and last part, we present the most important results of our studies on spatial networks. We first present an empirical study of 131 street patterns across the world, and propose a method to classify the patterns based on the geometrical shape of the blocks. We then present a cost-benefit analysis framework to understand the properties and growth of spatial networks. We introduce an iterative model that can explain the emergence of a hierarchical structure ('hubs and spokes') in growing spatial networks. Starting from the cost-benefit framework of this model, we finally show that the length, number of stations and ridership of subways and rail networks can be estimated knowing the area, population and wealth of the underlying region.

Throughout this thesis, we try to convey the idea that the complexity of cities is – almost paradoxically – better comprehended through simple

approaches. Looking for structure in data, trying to isolate the most important processes, building simple models and only keeping those which agree with data, constitute a universal method that is also relevant to the study of urban systems.

## PUBLICATIONS

---

Most of the ideas and figures presented in this thesis have appeared previously in the following publications:

- [1] Rémi Louf, Pablo Jensen, and Marc Barthelemy. Emergence of hierarchy in cost-driven growth of spatial networks. *Proceedings of the National Academy of Sciences, U.S.A.*, 110(22):8824–8829, 2013.
- [2] Rémi Louf and Marc Barthelemy. Modeling the polycentric transition of cities. *Physical Review Letters*, 111(19):198702, 2013.
- [3] Rémi Louf and Marc Barthelemy. How congestion shapes cities: from mobility patterns to scaling. *Scientific Reports*, 4:5561, 2014.
- [4] Rémi Louf, Camille Roth, and Marc Barthelemy. Scaling in transportation networks. *PLOS ONE*, page 0102007, 2014.
- [5] Rémi Louf and Marc Barthelemy. Scaling: lost in the smog. *Environment and Planning B: Planning and Design*, 41(5):767–769, 2014.
- [6] Rémi Louf and Marc Barthelemy. A typology of street patterns. *Journal of The Royal Society Interface*, 11(101):20140924, 2014.
- [7] Rémi Louf and Marc Barthelemy. Patterns of residential segregation. *Submitted*, 2015.

This thesis was the occasion to study other topics, that I chose not to address in the present dissertation. Some of this work appeared in the following publication:

- [8] Christian Borghesi, Laura Hernández, Rémi Louf, and Fabrice Caparros. Universal size effects for populations in group-outcome decision-making problems. *Physical Review E*, 88(6):062813, 2013.

Parts of the work presented here will also be the subject of a book that is being written concurrently with this dissertation

- [9] Marc Barthelemy and Rémi Louf. *Morphogenesis of Urban Networks*. Springer, 2016.



## ACKNOWLEDGEMENTS

---

*I had the idea that the world's so full of pain  
it must sometimes make a kind of singing.*

— Robert Hass, *Faint Music* [109]

I would first and foremost like to express my gratitude to Alain Barrat, Michael Batty, Pablo Jensen, Renaud Lambiotte, Jean-Pierre Nadal and Lena Sanders for accepting to be in my thesis committee. I can only imagine how daunting reading a PhD thesis must be, and I am really grateful for their time and for helping me to jump over this last hurdle. I sincerely hope this was not as awful as reading a 200 pages document (written by a student) sounds.

Thank you, also, to the Institut de Physique Théorique. By that, I mean its director Michel Bauer, the PhD coordinators Olivier Golinelli and Stéphane Nonnenmacher, and the incredibly efficient (and kind!) administrative and IT staff.

This thesis would not have been the same without the supervision of Marc Barthelemy. I still remember the sunny, winter morning of 2012 when the bus dropped me in the middle of nowhere, not far from the Institut de Physique Théorique. The fields were frozen, and vaguely white. I remember thinking, as I stepped out of the bus and felt the cold air biting my face, that this was going to be temporary. Needless to say, I did not sign up for three more years because of my love of long commutes, or the (too) familiar smell of the RER B. Neither was it for the office with a view, not on the fields—or the odd rabbit coming out of the nearby forest—, but on a brick wall. And faded flowers.

Now that this thesis is coming to end, I do know why I stayed, what I am grateful for. I am grateful to Marc for his trust, and for giving me the appreciated freedom to choose and explore the topics I liked. I am thankful, not only for his scientific supervision, our insightful discussions, but also for teaching me the things that cannot be found in textbooks: how to navigate the research world. Marc also knew when to push me and how to push me, and this is hopefully reflected in the work presented here. I am greatly indebted to him.

Je tiens à remercier mes parents, sans qui, plus encore que toutes les personnes sus-citées, cette thèse n'existerait pas. Souvent (!), on me demande: “Comment est-ce qu'un fils d'agriculteur fait pour devenir physicien?”. Je ne sais pas exactement, parce que c'est juste arrivé, et parce que cela a pris tant d'années. Je sais en revanche—c'est ce que

je réponds souvent—que c’est que ça n’aurait jamais été possible sans vous, et l’éducation que vous m’avez donnée. Vous m’avez offert le plus beau et le plus utile des cadeaux. Merci. Je vous en suis infiniment reconnaissant.

I finally wanted to thank Nikki, Clément, Pu, Rosa-Lỳ, Kévin (and my parents, again) for being there when it really mattered. And all my other friends and colleagues, for being there.



## CONTENTS

---

<b>I</b>	<b>INTRODUCTION</b>	<b>1</b>
1	STUDYING CITIES	3
1.1	We need data . . . . .	4
1.2	Cities as complex systems . . . . .	4
1.3	Layers and scales . . . . .	6
2	QUANTITATIVE REVOLUTION(S) IN URBAN SCIENCE	9
2.1	The first quantitative revolution . . . . .	9
2.2	A second quantitative revolution? . . . . .	11
3	METHODOLOGY	15
3.1	Of models and theories . . . . .	15
3.2	Quantitative stands for 'data' . . . . .	16
4	ABOUT THIS THESIS	21
4.1	Outline . . . . .	21
4.2	Miscellaneous notes . . . . .	22
<b>II</b>	<b>POLYCENTRI-CITY</b>	<b>25</b>
5	THE (END OF THE) MONOCENTRIC CITY	27
5.1	From monocentric to polycentric cities . . . . .	27
5.2	How to count centers . . . . .	32
5.3	The polycentric transition . . . . .	37
5.4	Summary . . . . .	40
6	HOW CONGESTION SHAPES CITIES	41
6.1	Fujita and Ogawa . . . . .	41
6.2	Problems with the Fujita and Ogawa model . . . . .	44
6.3	Modeling mobility patterns . . . . .	45
6.4	Monocentric to polycentric transition . . . . .	48
6.5	Number of centers . . . . .	50
6.6	Conclusion . . . . .	51
7	DISCUSSION	53
7.1	Questioning and extending the model . . . . .	53
7.2	Shadows in the empirical picture . . . . .	54
7.3	Summary . . . . .	58
<b>III</b>	<b>SCALING</b>	<b>61</b>
8	INTRODUCTION	63
8.1	Probing cities with scaling laws . . . . .	63
8.2	A brief history of allometric scaling and cities . . . . .	66
8.3	Summary . . . . .	72
9	FROM MOBILITY PATTERNS TO SCALING	75
9.1	A naive approach . . . . .	75
9.2	Beyond naive scalings: modeling the mobility patterns . . . . .	82

9.3	Discussion . . . . .	87
10	INTERPRETATIONS AND IMPLICATIONS OF SCALING LAWS	93
10.1	What scaling laws tell us about cities . . . . .	93
10.2	What cities? . . . . .	96
10.3	Conclusion and perspective . . . . .	100
IV	SEGREGATION	101
11	WHAT SEGREGATION IS NOT	103
11.1	Studying segregation . . . . .	103
11.2	Think first, measure later . . . . .	104
11.3	The dimensions of segregation . . . . .	104
11.4	The unsegregated city . . . . .	105
12	PATTERNS OF SEGREGATION	109
12.1	Introduction . . . . .	109
12.2	Presence of categories . . . . .	110
12.3	Measuring the attraction and repulsion of categories . . . . .	113
12.4	Emergent social classes . . . . .	115
12.5	Larger cities are richer . . . . .	117
12.6	Delineating neighbourhoods . . . . .	117
12.7	Poor centers, rich suburbs? . . . . .	124
12.8	Conclusion and perspective . . . . .	125
V	URBAN NETWORKS	129
13	A TYPOLOGY OF STREET PATTERNS	131
13.1	Streets versus blocks . . . . .	132
13.2	Characterizing blocks . . . . .	133
13.3	A typology of cities across the world . . . . .	136
13.4	A local analysis . . . . .	138
13.5	Discussion and perspectives . . . . .	138
14	COST-BENEFIT CONSIDERATIONS IN THE GROWTH OF SPATIAL NETWORKS	141
14.1	The model . . . . .	143
14.2	Crossover between star-graph and Minimum Spanning Tree	145
14.3	Spatial Hierarchy . . . . .	148
14.4	Efficiency . . . . .	153
14.5	Discussion . . . . .	156
15	SCALING IN TRANSPORTATION NETWORKS	161
15.1	Framework . . . . .	162
15.2	Subways . . . . .	163
15.3	Railway networks . . . . .	167
15.4	Summary . . . . .	170
VI	CONCLUSION	173
16	CONCLUSION	175
16.1	Lessons learned . . . . .	177
16.2	If I had to write a second thesis (Future directions) . . . . .	178

BIBLIOGRAPHY

181



## Part I

### INTRODUCTION

We begin this part with a general introduction that stresses the ever growing importance of cities in the world, and highlights the difficulties encountered when trying to reach a scientific understanding of these systems. We briefly outline the history of the quantitative tradition in the study of urban systems, and argue that we may be witnessing a second quantitative revolution. We then succinctly present the methodology that we followed during the past 3 years, and end this part with an outline of the content presented in this thesis.



## STUDYING CITIES

*Chaos was the law of nature;  
Order was the dream of man.*

— Henry Adams [10]

Cities appeared some 10,000 years ago [18, 163] concomitantly with the agriculture revolution, and really started to thrive after the industrial revolution [18]. In England first, where the revolution was born; London was the first city in the modern world to reach 1,000,000 inhabitants at the beginning of the 19th century. The urban growth then slowly spread through the end of the 19th and the 20th to the rest of the Western world. Now, while western countries are already mostly urban (as of 2014, the United States' population was 82% urban, Japan's cities hosted 93% of the population, and most countries in the European Union were around the 80% mark), most of what has been dubbed the 'urban revolution' is happening in developing countries. A symbolic barrier was reached in 2005, when it was estimated by the U.N. that more than 50% of the world total population was living in cities. It is not difficult to convince oneself that urbanisation is not an accident in human history, and that cities' influence and impact are not going to stop growing any time soon.

Source:  
UN Population  
Division (2011)

In fact, the impact of cities is already tremendous. First, they have a disproportionately large importance in the world's economy. A 2012 report by McKinsey noted that while cities represented respectively 79% and 19% of the United States' and India's population, their share in the countries' GDP was respectively 85% and 39%. Data from the NASA indicate that urban areas cover a total of 5% of the total land surface area in the world, roughly the equivalent of the superfcy of the European Union. Yet, despite their little spatial footprint, cities have a great impact on the environment. The United Nations indeed estimated in 2011 that cities were responsible for 70% percent of the world's CO<sub>2</sub> emissions.

We could multiply the statistics, but the few examples given above should convince the reader of the importance to understand cities if we want to improve the world we built for ourselves. The dramatic growth of urban areas in developing countries brings unprecedented challenges. The cause, and the solution of some of the world's most pressing challenges certainly find their origin in cities. By improving the way cities work, we can hopefully make dramatic changes to the way people live. To be able to do so however, we first need to understand how they work.

### 1.1 WE NEED DATA

Walk a few steps in your favourite city, feel the streets bustling all around you. The sound of the cars, of people chatting, the pavement lined with homogeneously diverse buildings. The sense of familiarity we feel when stepping back in a city that was once our home, years later. And that smell you had forgotten you knew. Maybe the hardest thing, when studying cities, is the impression that we know them closely. The belief that our impression of what they are, the way we experience them, gives a true picture of what they really are, the purpose they serve. This familiarity is what makes the study of macroscopic, human-made systems so difficult compared to the study of natural systems.

There are indeed only so many ways one can get acquainted with, say, electrons, and therefore just so many things one can say about them. This, in a sense, makes the study of electrons easy. Think about cities now. All the memories, habits, knowledge you have gathered over the years. As individuals, we know too many and too little things about them at the same time. We can have a very detailed recollection of the city we have experienced. But this information is not organised, and it is too local, too provincial. Therefore, we cannot infer what cities are solely from our own experience. We are a single piece of a puzzle that counts hundreds of thousands, millions of them, all with a different opinion of what their environment is like.

No, to understand cities, how they work as a system, we need to be told these thousands of stories, we need to analyse them and see how similar, or dissimilar they really are. To understand cities, we need data.

### 1.2 CITIES AS COMPLEX SYSTEMS

#### 1.2.1 *A paradigmatic example*

Cities are paradigmatic examples of complex systems [128]. First, they comprise thousands, millions of individuals that are moving and interacting constantly. Cities are indeed more than the mere agglomeration of residences, factories and shops in the same region; they exist and thrive through the resulting facilitated interaction between individuals [38, 207]. Cities are built so that many people can live together and interact.

Second, cities are incredibly resilient systems. There are multiple examples in History of cities that were completely destroyed – Dresden and Hiroshima, for instance, completely burnt to ashes during WWII – but were later rebuilt and thrived again.

Finally, cities exhibit very particular shapes and behaviours. Because of these identifiable properties, they are patterns that stand out in their environment [66]. We can recognise cities because of their particular



structure, even though the details of the structure differ from one city, country to another. The road network, for instance, is such that cities can be readily identified when looking at a map (even though the layout of say American cities is different from that of most European cities). The high density of population, hence nightlights, also make urban environments identifiable on satellite pictures. These are two obvious, visual particularities of cities, but some of their regularities are more subtle. In this thesis, we will be interested in some of these particular behaviours.

### 1.2.2 *An organised complexity*

The systems studied in Physics can be roughly divided in two categories [175]

- Simple systems with only a few variables. Their dynamics is described by *deterministic* equations. For instance, the motion of planets can be described with high accuracy by General Relativity.
- Weakly, locally interacting systems, with a very large number of particles. Their properties are described using *probabilistic* language. For instance, monoatomic gases in usual conditions of pressure and temperature are well described by Statistical Mechanics.

Cities, however, do not fit in any of the above categories. They are clearly not simple, deterministic systems, and cannot be described in their entirety with only a few variables. On the other hand, the traditional approach of Statistical Mechanics is also bound to fail. Although they can contain several million of individuals, cities are not maximally disordered systems, and thus cannot be described in the same way we describe gases. Cities, while being disorganised, have structure. Our goal is to identify and quantify this structure.

At the individual level, interactions are weak: one individual is very unlikely to radically change the system's dynamics. But the multiplication of individual interactions can create robust and influent structures (the activity centers discussed in Chapter ii, for instance). Interactions can occur locally – during face-to-face meetings – but also non-locally – through the phone, or the use of information systems. Individuals are not aimless particles, but usually have a purpose whenever they move. But at the same time, the sheer number of individuals leaves room for unexpected situations and encounters. As a result, cities are neither completely organised systems, nor are they completely disorganised. They are thus very different to the kind of systems natural sciences have traditionally studied.

### 1.3 LAYERS AND SCALES

A first step in the identification of order consists in identifying the different spatial and temporal scales involved in the dynamics within and of cities. The goal of any theory of how cities work would be to understand the phenomena occurring at each scale, to understand how scales interact with one another, and to establish a hierarchy of mechanisms, as in natural sciences [210].

#### 1.3.1 *Layers*

At the smallest scale, we have the individuals who live in urban systems. They make decisions about where they live, where they work, etc. and interact constantly with one another. Individuals are, in a way, the building blocks of cities, and it is therefore crucial to understand the way they interact with their environment to understand the structure and behaviour of cities.

At a larger scale, cities can be considered as systems characterised by specific behaviours [39]. Besides, they do not evolve in isolation and belong to larger scale structures. To quote the geographer B.J.L. Berry, ‘cities [are] systems within systems of cities’ [34], and their interactions—migrations, commodity and capital flows—ought to constrain their evolution [184].

Finally, there is a great amount of evidence to show that systems of cities also exhibit very particular behaviours: the rank-size plot of the population of cities that belong to these systems is indeed strikingly regular (a regularity known as ‘Zipf’s law’), and breaks down for other geographical units or when the chosen set of cities is not geographically and economically coherent [64].

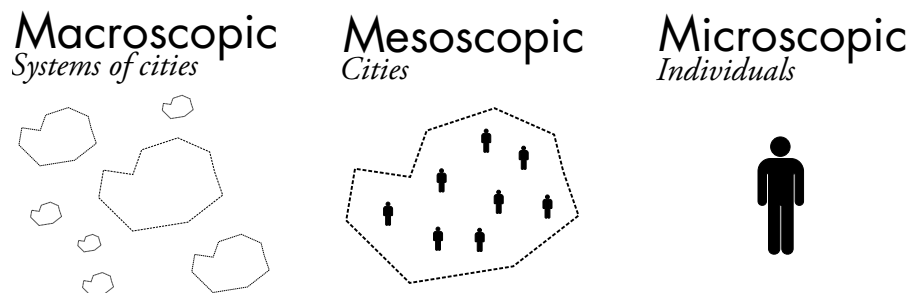


Figure 1: **Interactions at different spatial scales.** Cities are the result of interactions occurring at different spatial scales. The movement and interactions of individuals result in the properties of the city as a whole. But cities are not closed systems, and interact with other cities in a system of cities.

Cities are therefore the result of interactions occurring at different spatial scales. Furthermore, they are not static: they evolve in time, through various processes taking place at different time scales.

### 1.3.2 Time scales

First we have time scales of the order of a day, which span the daily commuting of inhabitants. This incessant movement of people has been traditionally explored through surveys, but new data now allow more thorough studies. The digital traces that are left by people at all times (through their mobile phone, metro pass or GPS device) indeed allow us to explore the structure of flows and the pace of life in cities at unprecedentedly fine spatial and time resolutions.

Then, at the order of a year one can see the variation in terms of wealth, population, etc. of cities, as recorded by statistical agencies. Data about demographic, social and economic aspects of urban systems allow us to characterise more specifically the structure and behaviour of these systems.

Finally, at time scales of the order of a decade, we can see the city's infrastructure as well as its spatial footprint evolve. The study of the underlying processes is made possible by various projects lead by the GIS community, historians and geographers which aim at digitizing historical maps of the road and rail networks in different regions of the world. Also, since the 1970s, many satellites have been taking pictures of the Earth's surface, and the remote sensing community has been treating these data to get information about the spatial extension of cities. These data should give us some insight about the processes responsible for the long-term evolution of cities' structure.

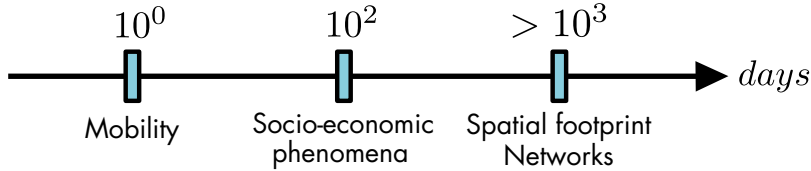


Figure 2: **Different time scales.** The various data available about cities are associated with different time scales.

These time scales are summarised on Fig. 2. The long-term goal of our studies is to understand exactly how cities and systems of cities behave, and how interactions between these three layers lead to the behaviours we observe.



## QUANTITATIVE REVOLUTION(S) IN URBAN SCIENCE

---

*And the first one now  
Will later be last  
For the times, they are a-changin'*

— Bob Dylan

It is difficult to make a concise summary of what is known and not known about urban systems. The vast amount of knowledge that has been gathered so far seems very little in comparison to the bewildering complexity of the object being studied [26]. Every map, every satellite view, every statistic, every step in cities elicits a question yet to be answered. What do we have to answer them? A surprisingly small array of empirical tools and models. A surprisingly small amount of solid, undisputed empirical facts.

Having said that, previous contributions are by no mean negligible. The body of quantitative knowledge about cities has dramatically grown since the quantitative revolution that took place in Geography after the 1950s.

People have recently suggested that we may be witnessing the dawn of a second quantitative revolution [28]. In the following Chapter, we will try to get some perspective on this claim, and see to what extent it is justified. We will start with a (very) brief account of the first quantitative revolution and the main themes around which it articulated knowledge (a more comprehensive account can be found in [200]). We will then critically review the factors usually invoked to justify the use of the expression 'second quantitative revolution'.

### 2.1 THE FIRST QUANTITATIVE REVOLUTION

Quantitative efforts in the study of human activities find their origin in Von Thünen's model of agricultural land in 1826. More than a century later, in 1933, the German geographer Walter Christaller published his Central Place Theory [58], which aimed at explaining the size and location of settlements in a system of cities. Needless to say, these early efforts are theoretical in nature, and the empirical aspect – studying things as they are – is left out. Likely because of the lack of available data.

The quantitative effort really starts to spread in the US in the 1950-1960 [36]. From the very beginning, the objective to make geography

a science is clearly stated, starting with the introduction of Bunge’s seminal *Theoretical Geography*, published in 1962 [52]. According to the author, geographers can and should go beyond the mere accumulation of facts, and try to discover the laws that rule the human and physical phenomena occurring on the Earth’s surface.

Bunge proposed geometry as a tool to understand the observed patterns and describe objectively the geographical space. The range of tools used quickly expanded [106, 56], spanning statistical models [125, 51] – whose importance is demonstrated by the publication in 1969 of Leslie King’s *Statistical Analysis in geography* – and graph theory – as early as 1963 with the publication of Kansky’s PhD thesis [121]. An early review of the use of graph theory in geography can be found in Hagget and Chorley’s book [107].

The research undertaken in the quantitative tradition can be – tentatively – divided in three different categories. First, the study of spatial differentiation aims at characterising the spatial patterns that result from human activities. For instance, the study of population or employment densities (see Part ii), the local concentration of population categories (see Part iv), or the repartition of cities inside a territory.

Second, the study of spatial interactions. The progressive realisation that distance is a critical factor to understand the arrangement of different spatial phenomena led Tobler to state the First Law of Geography [221].

Everything is related to everything else. But near things are more related than distant things.

Linked to the study of spatial interactions is the (in)famous gravity model, which states that the flow  $F_{ij}$  between two locations  $i$  and  $j$  is given by a function of the form

$$F_{ij} = C P_i^\alpha P_j^\beta f(d_{ij}) \quad (1)$$

where  $f$  is a decreasing function of distance. Although the analogy with Newton’s gravitation law was used by Reilly in 1931 to find the retail market boundaries between cities [192], the above formulation in terms of flows was formulated by Stewart in [215]. Note the competing existence of Stouffer’s theory of intervening opportunities [217], according to which the flow between  $i$  and  $j$  is proportional to the number of opportunities at  $j$  and inversely proportional to the number of opportunities between  $i$  and  $j$ . It was mathematically formulated much later by Simini et al. [208].

Finally, the study of infrastructure, which started with Kansky in 1963 [121]. The study of the shape and growth of road networks, railway networks and other infrastructure has recently witnessed a renewed interest thanks to the study of spatial networks [22].

## 2.2 A SECOND QUANTITATIVE REVOLUTION?

People can be forgiven for believing that the present time bears any sort of special character. But when we look closely enough, the change is perpetual, and what is new now will be outdated tomorrow. During the past 3 years, I have at many times overheard discussions about the fact that we were currently witnessing a 'second quantitative revolution' in the study of geographical systems. But is it really the case? What differences with past tools or methods could justify such a claim? In the following, we explore the three following hypotheses

- The new quantitative revolution is due to the use of new methods coming from interdisciplinary studies;
- The new quantitative revolution is due to the availability of 'new data';
- The new quantitative revolution is due to a technological convergence.

### 2.2.1 *New methods*

The recent years have seen the application of new methods, mainly coming from Physics or Computer Science, to the study of cities [73]. Either by geographers, or outsiders who imported well-established methods from another field [25]. These collaborations, or incursions, are however not new. For instance, John Stewart, an american astrophysicist is famous for the first use of allometric scaling in the study of cities [214], or for his work on the gravitation model [215]. Another interesting example is given by the collaboration in 1971 between Waldo Tobler – a geographer – and Leon Glass – a chemist – who plot the radial distribution function of Spanish cities, a method that is traditionally used to study the property of liquids [101].

So, the application of well-established methods from other fields to cities is not new, and neither are the contributions made by outsiders. Yet, we can identify two qualitative changes: the number, and nature of these contributions. If some authors have continued to import directly methods and models from other disciplines (for instance, the use of diffusion-limited aggregation models, traditionally studied in physics, to explain the growth of cities [144]), this type of theoretical contribution is becoming marginal. Contributions are more and more empirical; and if theoretical, are not direct applications of another domain's theories. For instance, Rozenfeld and co-authors used percolation on census tracts to define cities [196] in an original way. Masucci et al. use percolation on the road network for the same purpose [153], while Li et al. use percolation to study the properties of congestion [139]. New approaches to spatial network [22] have yielded new insights into the structure

and evolution of road, railway and subway networks [218, 24, 1, 4, 6]. Original out-of-equilibrium models that are inspired by the studied system allow a better understanding: Simini’s radiation model [208, 209] – which is nothing else than the mathematical transposition of Stouffer’s intervening opportunities theory – or our model to explain the polycentric transition of cities [2] are examples of such models. Not to forget the important literature on scaling relationships [39, 38, 3, 16, 5], and other empirical analyses – such as the study of residential segregation we present in Part iv.

At the same time, the number of contributions to the field from authors who do not have a geography (or economics, urbanism, etc. for that matter) affiliation seems to have increased over the past years. After all, I am a theoretical physicist by training, and this thesis is officially a Theoretical Physics thesis. So, if the contributions of outsiders are not new, they are changing in number and nature. To the point where we can wonder whether some of these ‘outsiders’ should still be considered as such.

### 2.2.2 *New data?*

Besides the import of methods from other disciplines, it is often argued that the influx of new data, thanks to the digitization of our lives, is a revolution in itself.

The most important new source of data come from the wide use of mobile phones across the world [102, 84]. They consist, for each individual, of a list of antenna locations to which the individual was the closest at a given time (either when she used the phone, or when she switched from an antenna to another). Naively, one could think that mobile phone data are better than census-based data: they give a *continuous* information about the flow of individuals within the city (and are not limited to commuting), they cover a larger part of the population (which is critical in developing countries: censuses are not widely used due to the costs involved, but mobile phones have a high penetration rate), and are more spatially precise than released census data in urban areas (see Figure 3 for a comparison between the smallest INSEE areal units, and mobile phone antennas in Paris). But one needs to be careful. If mobile phone data are fine to monitor aggregate quantities (such as origin-destination commuting matrices [133], to map population changes during the day [141], or year [68]), one should be careful with the study of individual trajectories (such as in the seminal [102, 211, 212]). Indeed, the fact that positions are recorded every time a call is made by the user – events with a powerlaw inter-event time [211] and probably correlated with locations – is likely to introduce an important bias in the obtained trajectories. Not mentioning the spatial sampling introduced by the fact that positions are attached to a finite number of antennas. Unfortunately, no study has looked at



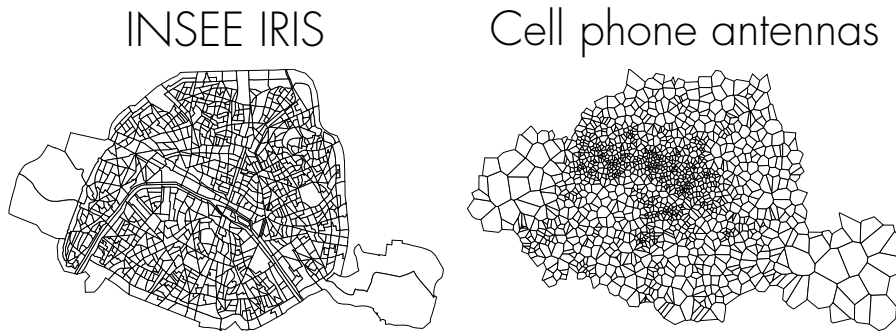


Figure 3: (Left) IRIS zones in Paris, the smallest statistical units defined by the national statistics institute, INSEE. (Right) Voronoi tessellation built from the position of antennas of a popular french mobile phone carrier. There are 40% more antennas than there are IRIS, and they tend to be more concentrated in zones of high daily activity (8th and 9th arrondissements).

the impact of these two types of sampling on the properties of the observed trajectories yet. In the meantime, one should refrain from using such data to study individual trajectories.

Mobile phone data are not the only ‘new’ source of data. Because mobile phones carry GPS chips that are used by applications such as FourSquare [167] or Twitter [134]. Last, but not least, credit card companies have recently started to release datasets regarding the spending of individuals [135].

So, new data (mainly mobile phone data) are now available and allow to give a picture of the city that was not accessible before. The contribution of these new data is particularly useful for the mobility of people besides commuting pattern [141], or for developing country where there are little census data available [43]. Are they so overwhelmingly different from previously available data to deserve the title of ‘revolution’? Nothing is less certain: in this thesis, for instance, I have only used traditional data sources, and we are still waiting for important results that ‘new data’ could teach us (and that we could not access with more traditional data). Only time will tell, and the term ‘revolution’ is not warranted yet.

### 2.2.3 A technological convergence

Interdisciplinary collaborations already existed, data were already there. So what is the qualitative difference between the state of the field say 20 years ago, and the state of the field as it is now, if any? A factor that is often overlooked is the recent technological leap in the treatment of information, including spatial information. Thanks to the development of GIS software as well as spatial databases and libraries, the treatment of geographical data has never been simpler. Added to this is the emer-

gence of powerful scripting languages, R and Python, which allow to quickly implement complex data analysis workflow or simulations, and reduce dramatically the time spent writing code.

Internet is also progressively changing the way research is done. Census data are more and more easily accessible available online. Open data repositories, although far from perfect, are emerging. Online platforms such as [www.github.com](https://www.github.com) allow to share and collaborate on code. All in all, the access and processing of information is getting easier and easier.

Taken individually, the introduction of methods from other disciplines, the increasing amount and specificity of available data and the technological progress in the treatment of information are probably not enough to justify the term ‘revolution’. Taken together, however, they could mark the beginning of a qualitative rupture in the way we understand cities.

It is too premature to conclude that the convergence of the aforementioned will necessarily deeply change our understanding of cities. Only the future can tell us whether new regularities, new laws are about to be discovered and more phenomena to be understood. But where there is data, there is hope. As long as the correct methodology is followed.

In the following Chapter, we will introduce the broad methodological principles that we adopted during this thesis.

*If it disagrees with experiment, it's wrong.  
In that simple statement is the key to Science.*

— Richard Feynman [86]

The success of natural sciences lies in their great emphasis on the role of quantifiable data and their interplay with models. Data and models are both necessary for the progress of our understanding: data generate stylized facts and put constraints on models. Models on the other hand are essential to comprehend the processes at play and how the system works. If either is missing, our understanding and explanation of a phenomenon are questionable. This issue is very general, and affects all scientific domains, including the study of cities.

Until recently, the field of urban economics essentially consisted in untested laws and theories, unjustified concepts that supersede empirical evidence [44]. Without empirical validation, it is not clear what these models teach us about cities. The tide has turned in recent years, however: the availability of data is increasing in size and specificity, which has led to the discovery of new stylized facts and opened the door to a new science of cities [28]. Yet, the situation is not perfect: while the recent deluge of data have triggered the apparition of many empirical analyses, in the absence of convincing models to explain these regularities, it is not always clear what we learn about cities.

In this chapter, we will try to specify what we mean by model, and explain with a concrete example why data analysis is not enough understand the behaviour of systems.

### 3.1 OF MODELS AND THEORIES

#### 3.1.1 *For what purpose?*

As scientific sceptics often like to remind us, all models, all theories are wrong. But surely, there must be some interest in models to make them deserve the months, sometimes years of work that scientist devote to them.

Models' two main functions are, broadly speaking, to understand, and to predict. The benefits linked with the ability to predict the behaviour of a system need not be recounted. Understanding is a more complicated notion, and a philosophical discussion of the concept lies

far beyond the scope of this thesis. Roughly, to understand is to untangle the mechanisms involved so as to have a simplified, barebone description of the processes that shape the system.

### 3.1.2 *Theory, not analogy*

Unfortunately, expressive words and metaphors are too often used as a substitute for a real understanding of the system. But, however intellectually appealing they are, metaphors are not a theory. For instance, what do we understand from the comparison of cities with biological systems? What new knowledge do we gain? Metaphors do not provide interesting ideas that are ready to be applied to a specific field. Rather, they trigger very different ideas into different people, which explains their recurrent success. Yet, what we need to highlight are regularities, not similarities.

We also need to avoid models that are only loosely connected to reality, analogy or metaphor. There is a lot of confusion, and little understanding to be gained that way. In the words of Einstein, Podolsky and Rosen

In a complete theory, there is an element corresponding to each element of reality. [79]

In this thesis, we tried to make sure that most – if not all – elements (variables) of our models are related to a quantity that is measurable. We also paid a special attention to the rigour in the language used. We qualify suggestions, by presenting them as such. This kind of work may be less suggestive, the vocabulary used less expressive, but it is a necessary step towards a science of cities. We need to clear the language of unfruitful metaphors and fill the gap with mechanisms.

## 3.2 QUANTITATIVE STANDS FOR 'DATA'

Richard Feynman's statement used as an epigraph in this chapter might be an oversimplified, narrow view of what Science is and how it proceeds. It nevertheless hits the nail right in the head, by isolating the core component of what Science is: a tight relation with empirical analysis. Data are needed, at first, to give us ideas about how the system works: stylized facts. We then usually try to build a simplified version of the system, a model, that is able to reproduce the stylized facts. Because of the simplification entailed, the model highlights the most important features of the phenomenon and allows us to understand the behaviour of the system. Finally, we use data again to test the predictions of the model and assess its validity and/or limitations.

In this thesis, we adopt a quantitative approach to the study of cities. In other words, we extract information about urban systems using measured quantities: data. As we will argue in the next section, however, data are not enough.

### 3.2.1 *Against data*

In 'Against Method', the philosopher of science Paul Feyerabend argued against the idea that Science proceeds through the application of a single, monolithic method; what people usually call 'The Scientific Method' [85]. The reference is not innocent, and I will argue here that, although empirical analysis constitutes the alpha and the omega of our enquiry for knowledge, data are not enough. There is common confusion, often innocent, that because data are at the core of scientific enquiry, one only needs data analysis to understand how a system works and predict its behaviour – especially so when we have a lot of data. A very extreme view of this statement has recently been put forth by Big Data supporters. An article in the magazine 'Wired' [14] recently argued that the current deluge of data marked the end of Science as we know it. That models were not necessary anymore, that they were to be replaced with the extensive correlation analysis that a vast amount of data allow. This view is completely misguided.

For one, pure data analysis is, at best, a myth: as Pierre Duhem argued in 1906 [71], all empirical observations are theory-laden. That is, they are necessarily affected by the theoretical presuppositions held by whoever is making the observation. Measuring the population of a city, for instance, presupposes that there are such objects as cities, and that we can delineate them. A deluge of data does not relieve the investigator from defining the objects she is studying, from implicitly thinking about the relation between the different elements in the system.

Then, correlations are science, indeed. But they are rudimentary science, and there is nothing new about them. Arguably, the reason why we are able to function at all as individuals is because our brain is capable of computing correlations all the time. Take chairs. Chairs are fairly simple objects. Yet, they come in all kind of colors, material and shapes. And despite this potentially infinite diversity, we are able to recognise a chair when we see one. We also have a notion of what a chair is to be used for. Although we do not acknowledge it often, we are capable of surprisingly high levels of abstraction and generalisation. Because our brains correlate, all the time.

Science starts with the observation of these regularities. For instance, that the sun always appears at the same place and disappears in the opposite directions. That seasons come and go regularly. That after the night always comes the day. Are pure correlations useful? Yes, for limited applications. Do they constitute science? No. Science is when one

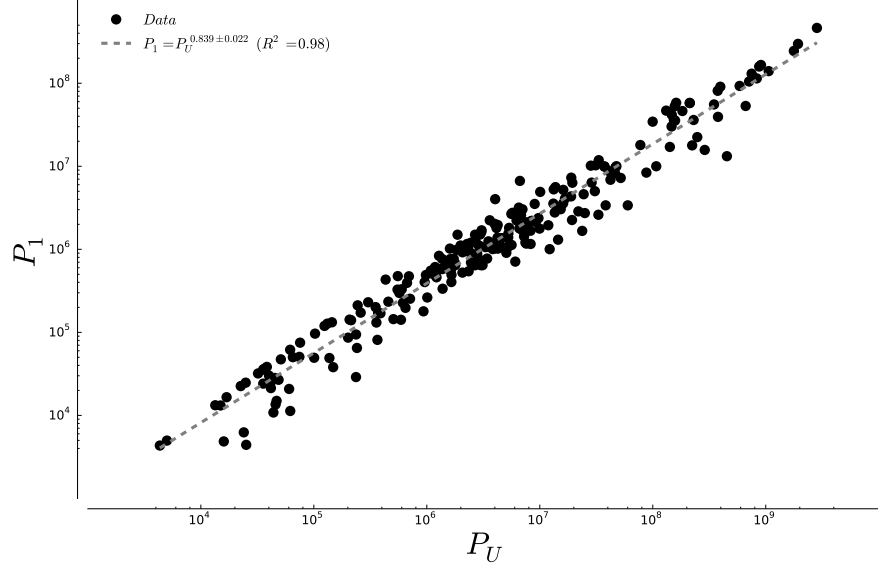


Figure 4: **The law of metropolises.** Population of the largest city of systems of cities  $P_1$  versus the total urban population  $P_u$  in that system. The dashed line shows the result of a powerlaw fit, whose exponent agrees well with the one found in [186]. Data for the total urban population and the population of the largest city of countries in the year 2000 were obtained from the World Bank.

goes beyond the simple observation of correlations, and tries to understand the mechanisms responsible for the correlations we observe.

In short, data is not enough: we must build models, theories.

### 3.2.2 An example: The law of metropolises

#### 3.2.2.1 Statement

The above discourse may seem a bit abstract, so let us observe the shortcomings of pure data analysis on a simple example, related to cities.

Using the GEOPOLIS database, Moriconi-Ebrard and Pumain derived a general transversal rule about system of cities, that they called *law of metropolises* [186]. If we note  $P_U$  the urban population of systems of cities (here countries), and  $P_1$  the size of their largest city, we can plot  $P_1$  versus  $P_U$  for all systems of cities and obtain the plot on Figure 4.

Assuming a powerlaw relationship between the two quantities, one finds

$$P_1 \sim P_U^{0.84} (r^2 = 0.98) \quad (2)$$

*The original regularity was observed for what the author calls 'metropolises', which are roughly equivalent to the largest city in terms of population.*

which agrees very well with the empirical data (for all years where data are available). It is tempting, at first, to consider this as yet-another empirical regularity exhibited by urban systems, and try to find a coherent interpretation in geographical terms. However, as we will show, if we assume that the Auerbach-Zipf law [17, 234] holds for each system of cities individually

1. We can derive a relation that fits the data as well as Eq. 2;
2. The relation is not a powerlaw.

### 3.2.2.2 Deriving the 'law of metropolises'

Let us consider a system of cities comprised of  $N$  cities, with total population  $P_U$ . The size of the largest city is noted  $P_1$ . We assume that the distribution of city sizes follows the Auerbach-Zipf law, so that the city of rank  $r$  (the  $r$ th largest city) has a population

$$P_r = P_1 r^{-\mu}$$

So the total population in the system of cities can be written

$$P_U = \sum_{r=1}^N P_r = P_1 \sum_{r=1}^N \frac{1}{r^\mu} \quad (3)$$

If we assume that  $\mu = 1$ ,  $P_U$  is given by the harmonic series, and thus

$$P_U = P_1 \left[ \ln(N) + \gamma + O\left(\frac{1}{N}\right) \right] \quad (4)$$

where  $\gamma \approx 2.58$  is Euler's constant. This gives us a first relation between  $P_1$ ,  $P_U$  and  $N$ .

Still using the assumption that the distribution of city size follows the Auerbach-Zipf law with  $\mu = 1$ , we can show (using extremal value theory) [60] that on average the size of the largest city is proportional to the total number of cities

'Average' as in  
ensemble  
average

$$P_1 \propto N$$

Thus, when the number of cities in the system is large,  $N \gg 1$  the following relation holds

$$\boxed{P_1 \ln(P_1) = P_U} \quad (5)$$

As one can see on Figure 5, the formula given by Eq. 5 fit the data as well as the previous one.

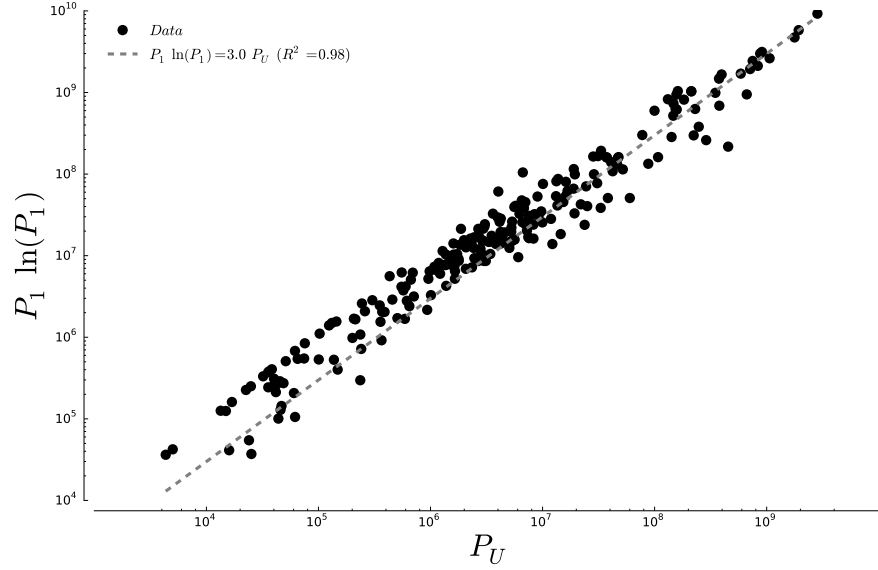


Figure 5: **The law of metropolises revisited.**  $P_1 \ln(P_1)$  versus the total urban population  $P_u$  in that system. The dashed line shows the result of a linear fit, which agrees as well with the data as does the powerlaw relation assumed in [186]. Data for the total urban population and the population of the largest city of countries in the year 2000 were obtained from the World Bank.

It is therefore impossible to determine which of Eq. 2 or Eq. 5 describes the ‘true’ relation between  $P_1$  and  $P_U$  based on data analysis alone. Nevertheless, the later finds a very simple explanation in the fact that cities in systems of cities follow the Zipf-Auerbach law up to a good approximation. In the absence of any theoretical explanation for the powerlaw relationship and given the empirical equivalence of both forms, it least-assuming to consider  $P_1 \ln P_1 \sim P_u$ .

### 3.2.2.3 Lessons learned

So, the *law of metropolises* is not a fundamental relation. This teaches us that, given the range of variation of the measured quantities, it is very difficult to distinguish empirically a powerlaw relationship from something qualitatively different such as  $Y \ln Y \sim P$ , as recently argued by Shalizi in [206]. One should therefore be wary of interpreting empirical relationships, like the one originally found in [186], unless a mechanistic explanation of the fitted relationship is provided. As shown above, what was thought as a fundamental law might end up being trivial and without great interest.

We will further discuss the limitations of data analysis in Chapter 10, after having studied scaling relationships.



## ABOUT THIS THESIS

*Anybody can plan weird, that's easy.*

— Charles Mingus

The following thesis might surprise the reader used to the monographs usually produced by PhD students in Social Sciences, articulated around a single, general question. The outline of this thesis reflects more the line of thoughts and of research that has been undertaken than the answer to a single question that would have been asked a priori and answered during the last three years. For that reason, the four Parts of this thesis are mostly independent. There is not single thread holding them together. But rather multiple wires; common themes and similar ideas.

## 4.1 OUTLINE

Part [ii](#) tackles the problem of measuring and understanding urban form, an issue that has been running through the 3 years of my PhD. In this Part, we first (Chapter [5](#)) present a brief historical overview of the monocentric and polycentric representations of the city, before enumerating the methods that are used in the literature to count the number of activity centers. We end with the observation that the number of activity centers increases in a regular way with population size. The following chapter (Chapter [6](#)) is devoted to an out-of-equilibrium model that we built in order to explain the previous empirical regularity. The model is able to predict the sublinear increase of the number of centers that we observe on American and Spanish data. In the last chapter (Chapter [7](#)), we question the assumptions of the model and the current empirical methods to quantify urban form.

Part [iii](#) is concerned with scaling relationships. We first propose (Chapter [8](#)) a non-exhaustive overview of the dawn and surge of allometric scalings, from Stewart's 1949 to the recent wealth of studies. Then, using the model developed in the preceding part, we show in Chapter [9](#) how the structure of mobility patterns allow us to understand the qualitative and quantitative values of the exponents related to urban form and mobility. We conclude this part with a discussion on the interpretation of these scaling laws, and their important shortcomings (Chapter [10](#)).

Part [iv](#) departs from the preceding chapters and turns to the study of residential segregation. Driven by the desire to extend the model

presented in Chapter 6, we soon realised there was a lack of robust empirical description of patterns of segregation that could be reproduced by a model. In Chapter 11 we tackle the problem of defining what segregation is; we propose a brief review of the existing literature, and subsequently define a null model – the segregated city. In the next chapter (Chapter 12), we build on this null model to propose a set of measures to quantify patterns of residential segregation.

Part v concerns the original topic of this thesis: spatial networks. Because my interests have shifted towards the study of socio-economical phenomena over the years, we only briefly present the most important results in the present thesis. The three chapters are, for the most part, reprints of articles that have been previously published in peer-reviewed journals. We first (Chapter 13) present an empirical study of 131 street patterns across the world where we propose a method to classify the patterns based on the geometrical shape of the blocks. In the following chapter (Chapter 14), we present a cost-benefit analysis framework to understand the properties and growth of spatial networks. We introduce an iterative model that can explain the emergence of a hierarchical structure (‘hubs and spokes’) in growing spatial networks. Starting from the cost-benefit framework of this model, we show that the length, number of stations and ridership of subways and rail networks can be estimated knowing the area, population and wealth of the underlying region.

Finally, Part vi ties everything together, highlights the lessons learned and concludes this thesis with some potentially interesting research avenues for the years to come.

## 4.2 MISCELLANEOUS NOTES

### 4.2.1 *Style*

I will be using the pronoun ‘we’ for most of the manuscript, to reflect the fact that the work presented here was, for the most part, done in the context of collaboration with others. For the sake of clarity, the technical details of calculations have been omitted in this manuscript. Most of these calculations are relatively simple anyway, and the interested reader can find them in the publications mentioned on page 2 of this thesis.

### 4.2.2 *Tools*

Unless otherwise specified, all figures in this manuscript have been prepared using Python 2.7 <sup>1</sup> and the Matplotlib library [112]. Inkscape <sup>2</sup>

<sup>1</sup> Available at <http://www.python.org>

<sup>2</sup> Available at <https://inkscape.org/en/>

was used to prepare most diagrams. This document was typeset using Vim and  $\text{\LaTeX}$ . The template used is the typographical look-and-feel `classicthesis` developed by André Miede.<sup>3</sup>

---

<sup>3</sup> Available at <http://code.google.com/p/classicthesis/>.



## Part II

### POLYCENTRI-CITY

The monocentric model of cities – where all activities are organised around a single activity center – has pervaded the literature on urban systems for more than 4 decades. However, as it was repeatedly demonstrated, the model is empirically inadequate.

The contribution of this part is threefold. First, we recount the history of ideas about urban form, from the monocentric hypothesis and its origins, to the various methods proposed to identify and count subcenters. We then demonstrate empirically the existence of a polycentric transition for cities, and that the number of centers increases as a sublinear function of population size. Finally, we propose an out-of-equilibrium model that explains the emergence of new subcenters as cities expand, and predicts the sublinear increase of the number of centers with population size.



## THE (END OF THE) MONOCENTRIC CITY

*It may be a small irony that just as  
the phenomenon of polycentricity  
is getting considerable attention,  
The world is moving beyond it.*

— Peter Gordon & Harry Richardson [103]

The hypothesis that cities organise themselves around a single center of activities – often called Central Business District (CBD) in the US – may well be one of the strongest hypotheses in urban studies. Although no one seriously believes in its validity anymore, its influence is still noticeable in many empirical and theoretical works. In order to deconstruct the monocentric model, we first need to understand where it came from in the first place, why it was introduced, and what evidence it was based on.

In this chapter, we present a historical perspective on the monocentric hypothesis. First, the context in which it was introduced, how it was gradually realised that cities had a decentralised structure, and the emergence of the notion of center. We then present a brief review of the methods and tools developed to count their number. Finally, using American and Spanish data, we show that larger cities are more polycentric. This suggests the existence of a transition from a monocentric to a polycentric structure when the population of cities increases.

## 5.1 FROM MONOCENTRIC TO POLYCENTRIC CITIES

Maybe the least assuming way to represent the density profiles in cities is through either choropleth maps, or 3-dimensional representations. On choropleth maps, the  $x$  and  $y$  coordinates correspond to the original coordinates projected on the plane. In the former case, the different values of density are expressed by the use of different colours. This approach can be traced back as far as 1898 in Meuriot's *Des agglomérations urbaines dans l'Europe contemporaine* [159] who drew a large number of density maps of large European cities. He was later followed by Jefferson in 1909 [117] who did the same for several cities in the US, Europe and Australia.

3-dimensional representations, on the other hand, use the  $z$  coordinate to represent the density values. On Figure 6 we represent the density profiles of two metropolitan areas in the US: Minneapolis-

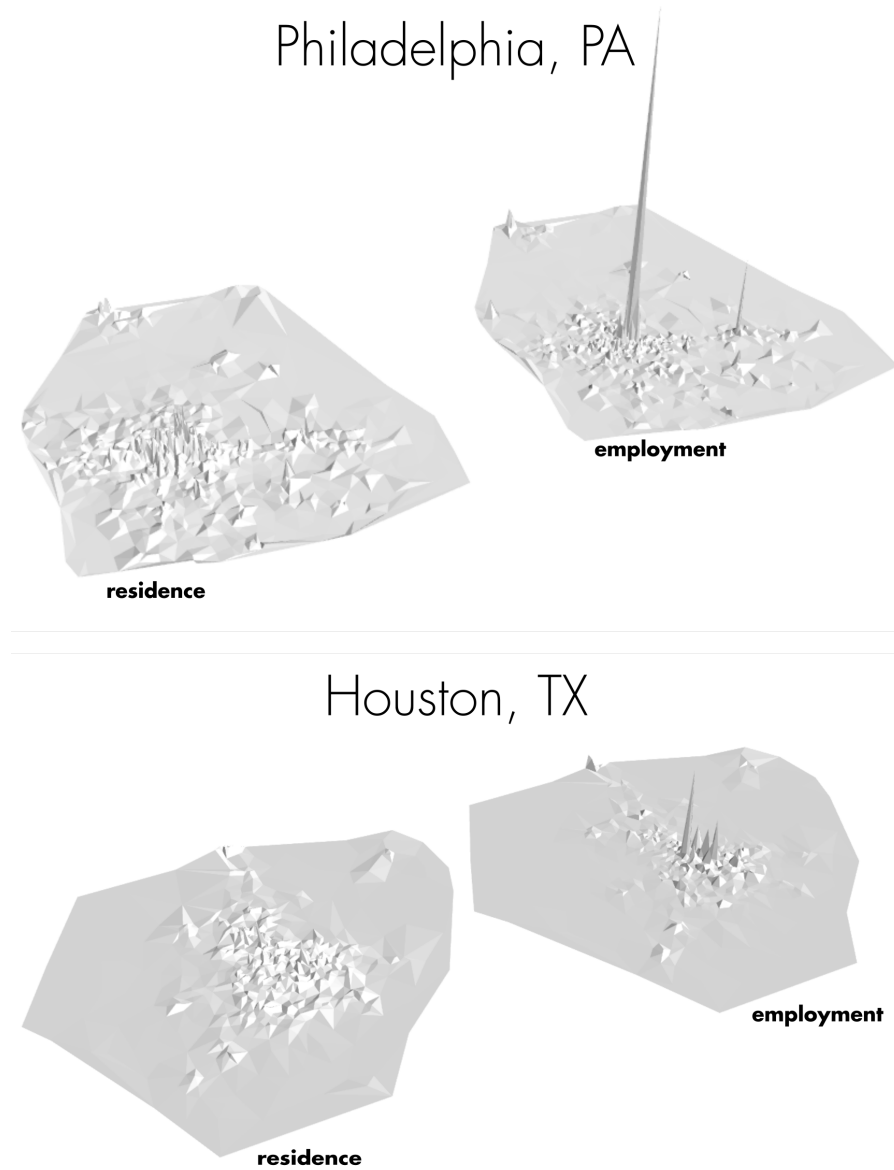


Figure 6: **3D representations of densities.** Residential and employment densities in (Top) the Metropolitan Statistical Area (MSA) of Philadelphia, PA and (Bottom) the MSA of Houston, TX. Employment and residential densities are represented at the same scale. Employment densities are sensibly more peaked than residential densities, suggesting that the notion of ‘center’ is more relevant in the context of activities. Data were obtained from the 2000 US Census.



St. Paul, MN and Houston, TX. These two cities are enough to illustrate the difficulties associated with studying density profiles.

*What densities we are talking about?* People are constantly moving throughout the city during the day, and density profiles can only be (approximate) snapshots of the city at different instants. Traditionally, scholars have only considered residence densities (nighttime city) and employment densities (daytime city). The recent availability of mobile phone data may however give us a more precise, continuous picture of the densities during the day [141]. In this part, we will be focusing on employment densities.

*How can we make sense of these density patterns?* The densities represented on Figure 6 are indeed very complex, and we would like to isolate some particular structure. Arguably, the notion of center stems from this desire to find some structure in the complex, messy empirical reality.

Realising that districts of large population tend to be central, and districts of small population in the periphery, Clark proposes in 1951 [59] to write the density  $\rho$  as a function of the distance  $d$  from the center

$$\rho = a e^{-d/b} \quad (6)$$

Where  $a$  is the density at the center, and  $b$  the typical distance over which the density decreases. To justify his assumption, Clark plots the population density of various cities as a function of the distance to the center [59]. Some structure was found. The monocentric hypothesis was born.

Looking at the density profiles plotted by Clark in 1951 [59] for many cities across the world, or on Figure 7 for the Minneapolis-St. Paul MSA, one can be forgiven for thinking that cities have a monocentric structure. Such profiles indeed almost always exhibit a sharp decrease as we go farther from the city center – defined here as the areal unit with the highest density.

However, density profiles are not enough to prove the existence of a monocentric structure. Unless one other hypothesis is verified: namely that the pattern of employment densities is symmetric under rotations around the center. This is however never the case: cities are nowhere isotropic but in the imagination of modelers. To make this point clearer, we show on Figure 7 both the density profile of the Minneapolis-St. Paul MSA and a map where we highlight in black the tracts with an employment density greater than  $10000 \text{ km}^{-2}$ . As one can see, two tracts (respectively the historical centers of Minneapolis, and of St. Paul) are highlighted. However, the peak in density corresponding to St. Paul is not distinguishable on the density profile. Indeed, it is averaged out with smaller densities that are located at equidistance from Minneapolis. The

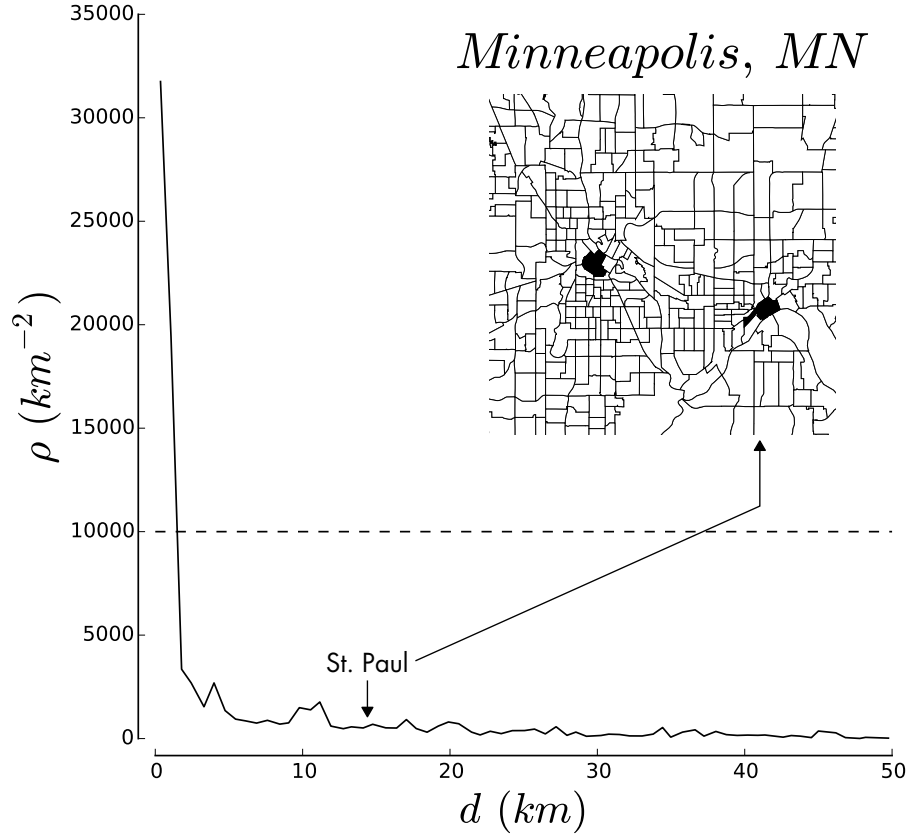


Figure 7: **The limitations of density profiles.** Employment density as a function of distance to the center for the Minneapolis-St. Paul MSA in 2000. The center is defined here as the tract with the highest employment density, and corresponds to the historical Central Business District of Minneapolis. The curve exhibits a very sharp decay, giving the illusion of a monocentric structure. (Inset) The census tracts of Minneapolis-St. Paul in 2000. In black, the census tracts where the employment density reaches values above  $10,000 \text{ km}^{-2}$ . The two tracts coincide with the historical centers of the Twin Cities, and are distant from 14 km. This fragmented structure cannot be inferred from the density profile (arrow on the curve).

decreasing exponential model, however appealing, is thus misspecified.

So why did Clark's methods and plots did not become a simple curiosity, but were instead so widely adopted? Although it is sometimes difficult to trace back the reasons for the adoption of ideas, there is little doubt that the echo this idea had in urban economics had something to do with it (besides the simplicity of the hypothesis). Indeed, beginning as an implied assumption in Clark's empirical analysis, the monocentric hypothesis first became clearly stated in the theoretical work of economists.

The Alonso-Muth-Mills model (inspired by Von Thünen's land rent model) might well be the reason for the long-lasting influence of the

monocentric model<sup>1</sup>. In 1964, Alonso introduced the bid-rent curve as a function of the distance to the city center [12]. The assumption that all firms in a city are concentrated in a single, fixed-size part of the city naturally followed. Later, in 1967 and 1969, Mills [161] and Muth [164] show how we can obtain an exponentially decreasing function for the density as a function of the distance from the center, using the monocentric hypothesis. The monocentric Alonso-Muth-Mills (AMM) model was born, and was seemingly backed by empirical evidence.

One should not underestimate how the monocentric model influenced people's perception of what a city is. In the US, the name of Central Business District is casually used as a way to designate the principle activity center in a city. Many, if not most, measures of the spatial variation of quantities inside cities actually use the notion of 'distance to the city center'. Many authors are relying on the monocentric hypothesis for their empirical analysis – sometimes without being aware of it. This bias can still be found in the recent literature. For instance, in a recent study by Glaeser, Kahn and Rappaport on the repartition of income classes in cities [100], the authors comment on plots of the average income as a function of the distance to the center. This only makes sense, however, under the assumption of monocentricity.

This persistence of the monocentric hypothesis is all the more surprising that authors repeatedly suggested and showed that the hypothesis was not adequate. In 1974, Kemper and Schmenner [124] explore industry and employment density data, trying to fit a negative exponential function. Their conclusion is clear: "A declining exponential function fails to explain much of the spatial variation of manufacturing density". A few years later, Odland [168] explores the possibility of polycentric cities on a theoretical basis. As explained in [104], scholars subsequently started to explore the density patterns of cities by fitting multi-center exponential functions of the form

$$\rho_i = \sum_{j=1}^q A_j e^{-d_{ij}/b_j} \quad (7)$$

where  $\rho_i$  is the density at location  $i$ ,  $q$  the number of centers,  $A_j$  the local maximum of density at  $j$ ,  $b_j$  the characteristic size of the center  $j$ , and  $d_{ij}$  the distance between locations  $i$  and  $j$ . The idea of polycentricity, originally as the generalisation of the monocentric hypothesis, is progressively gaining ground.

Trying to fit equations like Eq. 7 is cumbersome, and requires some a-priori knowledge of the density patterns. It requires to determine in advance which parts of the cities are going to be subcenters, before attempting to fit the density profile. As noted in [98], authors used arbitrary definitions of subcenters, either designating them based on their

**subcenter** because they are subsidiary to the traditional CBD

<sup>1</sup> A concise exposition of the AMM model can be found in [49, 90]

own intuition, or referring to the centers defined by planning agencies. The centers were thus determined *exogenously*.

In this context, the first definitions of employment centers independent from the exponential model start to emerge, and subcenters start an existence of their own. By the 90s, the idea that cities can be polycentric is well-established, and more and more empirical analyses confirm the existence of several employment centers. For instance, McDonald [155] identifies the employment subcenters in the region of Chicago, IL; Giuliano and Small [98] in the region of Los Angeles, CA; Dokmeci et al. [70] show that Istanbul's employment is spread across several centers, etc.

The concept of subcenters is further expanded in 1991 [95], when Garreau shows that secondary centers are not necessarily 'subcenters'. Indeed, activities do not always accumulate in the traditional downtown. He introduces the concept of 'Edge cities': the concentration of business, shopping and entertainment at the outskirts of cities, in regions that were previously rural, or purely residential.

## 5.2 HOW TO COUNT CENTERS

The methods designed to identify employment subcenters can be divided in three categories. The clustering methods, which appeared first, were progressively abandoned for regression-based methods due to their reliance on arbitrary cut-offs. Distribution-based methods have emerged recently, and leave aside the spatial aspect of the density distribution.

### 5.2.0.1 *Clustering methods*

In 1987, McDonald [155] remarks that despite being mentioned in the empirical and theoretical literature, the features that an employment subcenter should have are nowhere discussed. For the first time, he proposes a method to determine the number of subcenters empirically. Given a number  $T$  of areal units, we will say that  $i$  with employment  $E_i$ , population  $P_i$  and surface area  $A_i$  is an employment subcenter if:

*He also proposes a definition based on the employment-to-population ratio*

The gross employment density  $\rho_i = E_i/A_i$  is greater than that of the contiguous units;

Giuliano and Small [98] acknowledge the necessity to consider employment densities to define subcenters put forward by McDonald [155]. However, they deplore that the method does not allow for adjacent units with a high employment density to be centers – as only the larger one would be selected. Thus, they propose an alternate definition. Namely that a contiguous set of units  $\mathcal{S}$  is a subcenter if

- The employment density  $\rho$  of every areal units in the set  $\mathcal{S}$  is greater than a threshold value  $\bar{D}$ ;
- And the total employment  $E$  contained in  $\mathcal{S}$  is greater than a threshold  $\bar{E}$ .

where the thresholds  $\bar{D}$  and  $\bar{E}$  are imposed arbitrarily. Using this definition, all areal units with a high employment densities are part of a subcenter, unless they are small (contain less than  $\bar{E}$  employees) or isolated (i.e. they do not belong to a cluster containing at least  $\bar{E}$  employees).

As mentioned by Anas et al. in [13], because density landscapes are highly irregular at a small scale (see Figure 6 for instance), the subcenter boundaries are very sensitive to the threshold values. Because there is no a priori reason to choose a threshold rather than another, the obtained subcenter boundaries are arbitrary and may vary from one author, one situation to another. Instead, it would be preferable to have a method based on first principles, that adapts to the local specificities. In McMillen's words, threshold methods lack a proper consideration of how large is 'large' supposed to mean for the threshold values [157].

Another problem highlighted in [13] is that the number of centers depends on the size of the areal unit, an issue that is tied to scale problem discussed in the Modifiable Areal Unit Problem (MAUP) [172] literature. On the one hand, small areal units will lead to several low employment density units in otherwise very high density areas. On the other hand, large areal units are likely to smooth over local employment peaks. This begs the question of whether we should use contiguity of units, or rather distance, as a measure of proximity.

#### 5.2.0.2 Regression-based methods

In an attempt to address these concerns, McMillen [156] proposes a two stage procedure. In the first allegedly non-parametric stage, he uses a geographically weighted regression (GRW, see [51] for more details on the topic) to 'smooth' the employment density, using distance rather than contiguity as a measure of proximity, thus partially solving the issue linked with the size of areal units. The units that have unusually high employment densities compared to the broad spatial trends obtained with the GWR are designated as candidate subcenters. If we note  $\rho_i$  the employment density at site  $i$ ,  $\hat{\rho}_i$  the density estimated with GWR and  $\hat{\sigma}_i$  the standard deviation around this estimate,  $i$  is said to be a *candidate* subcenter if

$$\rho_i - \hat{\rho}_i > 1.96 \hat{\sigma}_i$$

Candidate, because the GWR only identifies fluctuations in the density profile with no consideration of whether these local fluctuations

have a sensible impact on the employment density. Identifying which of these candidates are actually centers is the goal of the second, semi-parametric procedure. This second procedure uses somewhat arbitrary criteria (the first and second largest candidates are omitted in the regression, candidates at less than 1 mile from the CBD are omitted) to produce a second reference global trend, to which real values are compared to identify the ‘real’ centers among the candidates.

Redfearn criticizes the first procedure [191], on the ground that candidate subcenters are defined as outliers with respect to an average that uses half of the total number of points (in the GWR), thus losing the local information about employment density. The author proposes another non-parametric method that aims at correcting the issues with McMillen’s [191]. The estimation of the employment density is done locally in order to keep intact the local structure of the density profile. However, arbitrariness still lies in the choice of the span (the amount of data that are considered to estimate the slopes at a given point) for the GRW. In other words, regression-based methods are *not* truly non-parametric.

#### 5.2.0.3 *Distribution-based methods*

The approach that we originally took in this thesis is radically different from that of regression-based methods [2]. We start with the remark that one does not need to know the spatial arrangement of areal units with different densities in order to know which ones are most important. Indeed, the local fluctuations that are registered as centers in the regression-based methods are very likely to have a negligible contribution to the total employment. They can thus be left out in a first approximation. A good estimate of the number of centers should thus be given by the shape of the employment density distribution alone. Because it does not require any spatial knowledge, it makes the extraction of centers fairly easy and quick to compute compared to the previous methods.

We start by building the rank plots of employment density  $\rho$  inside the areal units (see Figure 8). These plots display a decay at least as fast as that of an exponential. If they were an exact exponential, they could be modeled by a function of the form

$$\rho(r) = \rho_0 e^{-r/r_c} \quad (8)$$

where  $\rho(r)$  is the  $r$ th highest value of the density inside the city,  $\rho_0$  the maximum density value. This exponential decrease implies that there exists a natural scale for the rank,  $r_c$ , that we interpret here as the number of centers. In order to get the number of centers, one would

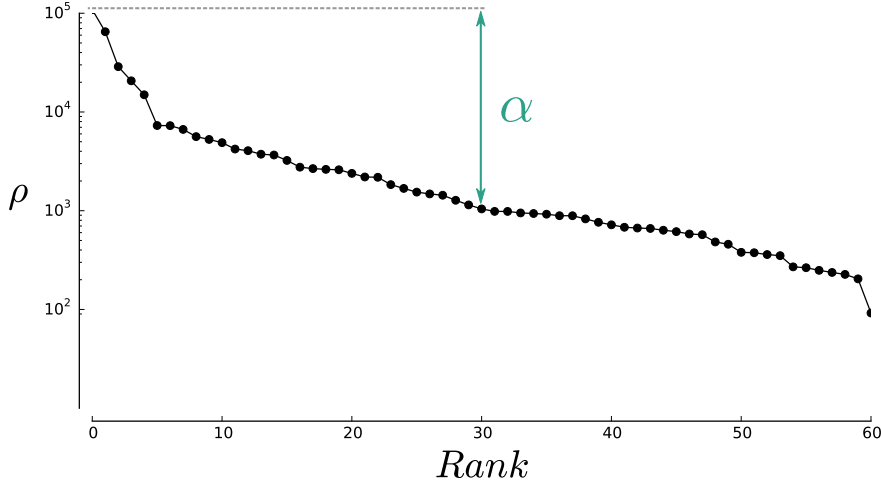


Figure 8: **The rank-plot method.** Rank plot of the employment density in the Zip Code Tabulation Areas of Los Angeles, CA.

either need to compute the slope on a lin-log plot, or find the value of  $r^*$  for which

$$\rho(r^*) = \frac{\rho_0}{e} \quad (9)$$

in which case  $r^* = r_c$ . However, empirical rank plots are not strictly exponential, and we define the number of centers using a threshold value  $\alpha$ . We define  $\rho_m$  as

$$\rho_m = \frac{\rho_0}{\alpha} \quad (10)$$

and the number of subcenters  $k$  is equal to the number of values  $\rho_c$  of the density such that  $\rho_c \in [\rho_m, \rho_0]$ .

In the case where the rank plot would be strictly exponential, we would have

$$k = \rho_0 \ln \alpha \quad (11)$$

so that the number of centers is mainly determined by  $\rho_0$ . Small variations in  $\alpha$  should not sensibly change the number  $k$  of centers obtained.

The method however suffers from two flaws. First, the use of an arbitrary parameter, the threshold  $\alpha$  to extract the number of centers. All the criticisms listed earlier also apply: we are not sure to extract the ‘true’ number of centers. Moreover, the method assumes a particular form for the density distribution, which is likely to bias the estimation.

Louail and Barthélemy [141] propose a generalisation of the previous method based on the Lorenz curve. Given the ordered set of densities

*The Lorenz curve is often used in Economics to quantify income inequality.*

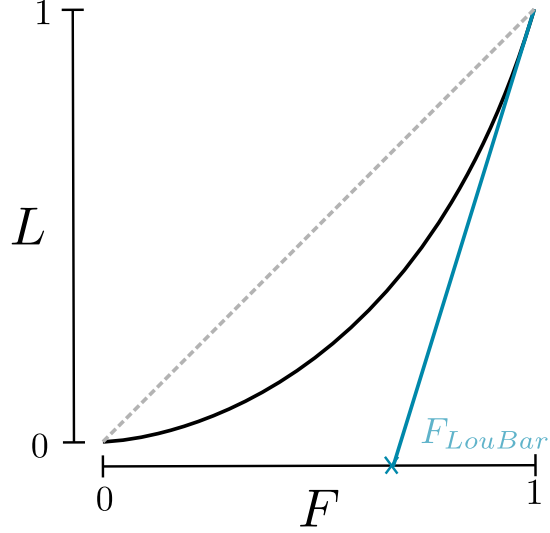


Figure 9: **Lorentz curve and Loubar method.** An example of realistic Lorentz curve (solid black line), the curve that would be obtained in a city with uniformly distributed density (dashed grey line), and the tangent at the point  $L(F) = 1$  (blue line) used to determine the number of centers in the LouBar method.

$\rho_1 < \rho_2 < \dots < \rho_T$  in the  $T$  units, we plot the proportion of cells  $F_i = i/T$  as a function of the corresponding proportion of employment density

$$L_i = \frac{\sum_{n=1}^i \rho_n}{\sum_{n=1}^T \rho_n} \quad (12)$$

so that both  $F_i$  and  $L_i$  take their values between 0 and 1 (see Figure 9). It is easy to see that, in the case of a city with a uniform employment density, the Lorentz curve is a straight line. In the general case, however, the curve has a convex shape, with a more or less pronounced curvature. The higher the curvature of the Lorentz curve, the higher the inequality in terms of employment density, and thus the smaller the number of potential centers.

Following this observation, the authors define a new criterion to determine the number of centers. They consider the intersection  $F^*$  between the tangent of the Lorentz curve at the point  $L(F) = 1$  and the axis  $F = 0$  (see Figure 9). The units that correspond to the values of  $F$  between  $F^*$  and 1 are defined as centers. This definition has the merit to only depend on the distribution of density inside the areal units; it is *genuinely non-parametric*, while being easily tractable and understandable.

Of course, all the methods presented here have issues (that we discuss in Chapter 7), and there is currently no consensus on what method should be used to find the employment centers. More work is needed



before we arrive at a satisfactory description of urban form. Nevertheless, the results given by these methods – although slightly different – provide together a compelling evidence for the polycentric structure of cities.

### 5.3 THE POLYCENTRIC TRANSITION

Occasionally mentioned in the empirical literature [157, 191], and hinted at in urban economics models [91], the greater polycentricity of larger cities was not firmly established before this thesis. Almost all cities (apart from the notable exception of twin cities) start growing around a single center of activity. Yet, as we will see, no large city adopts a strict monocentric structure. Therefore, it seems that, as they grow and expand, urban systems develop a more and more polycentric form. We call this phenomenon the ‘polycentric transition’ of cities.

#### 5.3.1 *Empirical evidence*

##### 5.3.1.1 *American cities (Census data)*

Historical data over long periods of time, on a consistent set of areal units, are very difficult – if not impossible to find. However, we do, for one point in time, have many cities with very different population values. We can thus compute and plot the number of centers as a function of population. Of course, as we will discuss in more details in Part iii, there is a gap between time series and transversal studies that is not completely obvious to bridge. Some cities can be, for historical reasons, locked into a monocentric state when the average city would not. For different reasons, another city might as well have developed a polycentric structure more pronounced than other cities of the same size have. The idea here is to look at a large number of cities and measure the average behaviour of this ensemble of cities, hoping that marginal cases are indeed marginal.

##### 5.3.1.2 *American cities (census data)*

During this thesis [2], we used data on the employment in the Zip Codes of US cities every year between 1994 and 2010. We first extracted the number of centers for every city, for every year between 1994 and 2010. Using the rank-plot method described earlier. We then applied the following treatment to the data:

- If there is only one Zip Code in the given city,  $k = 1$ ;
- We perform a Kolmogorov-Smirnov test [151] between the distributions of a given city for consecutive years. If there is a significant difference (above a threshold  $p_{KS}$ ) between the distribution

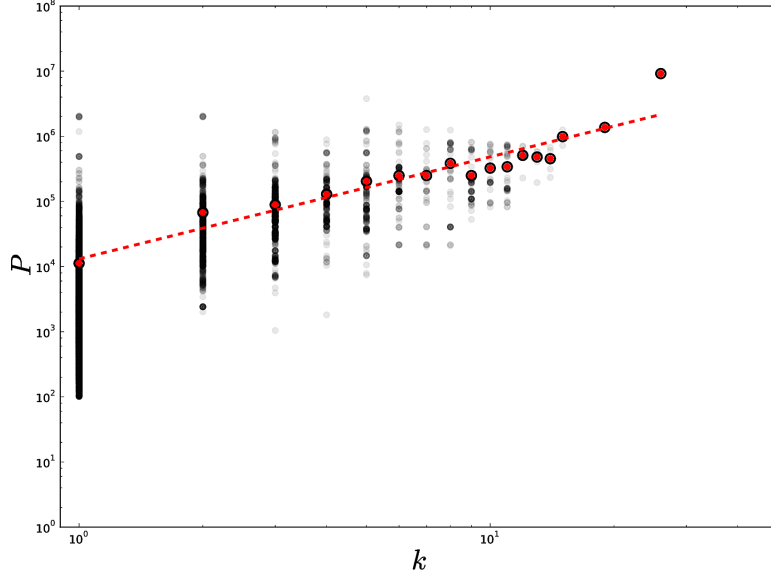


Figure 10: **Centers in American cities.** Scatter plot for the estimated number of centers versus the population for about 9000 cities (different realisations) in the US. The red dots represent the average population for a given number of subcenters. We fit this average assuming a power-law dependence giving an exponent  $\delta = 1.56 \pm 0.15$  ( $R^2 = 0.87$ ). Data were obtained from the US Census Bureau's Zip Code Business Patterns for every year between 1994 and 2010.

at  $t$  and  $t + 1$ , we keep the point at  $t + 1$ . If there is no sensible difference, we discard it.

At the end of this process, we obtain points that can be understood as coming from different realisations of a city. We then plot the number of centers computed for all these realisations as a function of the total population and obtain the curve obtained on Figure 10.

A power-law fit on the average per population bin gives an exponent  $\delta = 1.56 \pm 0.15$  (95% C.I.). Thus, we find that on average, the number of centers in US cities scales with population size as

$$k_{US} \sim P^{0.64} \quad (13)$$

#### 5.3.1.3 Spanish cities (mobile phone data)

Using mobile phone data and the LouBar method to determine the number of centers, Louail et al. [141] also computed the number of centers versus population for Spanish cities.

$$k_{Spain} \sim P^{0.64} \quad (14)$$

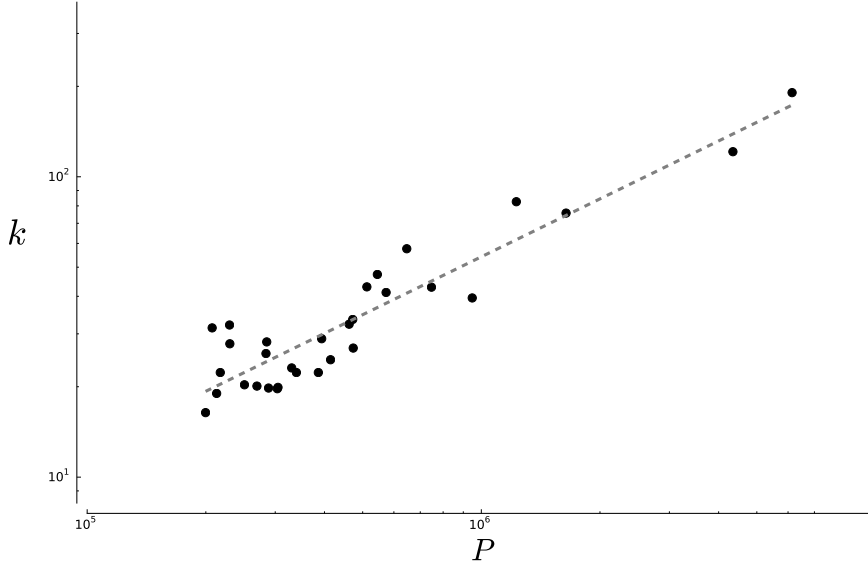


Figure 11: **Centers in Spanish cities.** Scaling of the number of centers with population for Spanish metropolitan areas. Assuming a powerlaw relationship, the authors of [141] find an exponent  $\beta = 0.64$  ( $r^2 = 0.93$ ). The data were kindly provided by Thomas Louail.

Strikingly, the exponent they found is very close (equal) to the one we found on a different system of city, using a different method to count centers, and a radically different data collection method.

Taken together, the previous empirical analyses teach us that

- The larger cities are, the more polycentric they tend to be;
- The average behaviour is well-approximated by a power-law relationship between the number of centers and population;
- The increase of the number of centers with population is *sublinear*.

These facts for a theoretical explanation. We will present a model to that effect in the next chapter. But before concluding, let us review quickly the reasons that are traditionally invoked for the polycentric transition.

### 5.3.2 Reasons invoked for the polycentric transition

There are numerous examples where polycentrism finds its origin in the fusion of two Metropolises, or the incorporation of satellite municipalities [131]. The Twin Cities in the US, for instance: the cities of Minneapolis and St. Paul have grown to such an extent that they now form a single metropolitan area. The region of the Ruhr in Germany, or the region of Tokyo in Japan are other examples. However, in this thesis, we are only interested in an endogeneous polycentrism, caused

by the growth of a single city.

Already in 1972, Mills [162] suggests that congestion might be the cause of decentralisation and suburbanisation in large metropolitan areas. However, we have to wait until 2003 for McMillen to propose a thorough empirical investigation [157]. Commuting cost is estimated using the peak travel time index which is defined as the ratio between the average travel time at peak congestion time over the average travel time at any other time of the day. Effectively, the commuting cost is thus a measure of the level of congestion in the city.

Studying US cities, The author finds a positive correlation between the number of centers, population, and commuting cost. In other words, congestion might be the key factor to understand the polycentric transition of cities.

#### 5.4 SUMMARY

In this chapter, we have presented a historical perspective on the monocentric hypothesis, trying to show why it appeared, disappeared, and how it is still hiding in some of the empirical literature. We then discussed the polycentric hypothesis, how it was introduced, and the different methods that have been proposed to identify and count the sub-centers.

We then showed on US and Spanish data that the average number of activity centers increases sublinearly with population size. This proves, we believe, the existence of a polycentric transition of urban areas as their population increases. A transition, we saw, that might be due to increased levels of congestion in larger cities. In the next chapter, we will present a model to understand this polycentric transition.

## HOW CONGESTION SHAPES CITIES

*What is here required is a new kind of statistical mechanics,  
in which we renounce exact knowledge not  
of the state of the system but  
of the nature of the system itself.*

— Freeman J. Dyson [77]

We saw in chapter 5 that as cities grow and expand, they evolve from a monocentric organisation where all the activities are concentrated in the same geographical area – usually the central business district – to a more distributed, polycentric organisation. In this chapter, we will try to uncover the mechanisms at play behind this transition. We begin with a brief introduction of the model of Fujita and Ogawa in urban economics. We will highlight its shortcomings, and present a stochastic, out-of-equilibrium model. This model relies on the assumption that the polycentric structure of large cities might find its origin in congestion, irrespective of the particular local economic details. We are able to reproduce many stylized facts, and – most importantly – to derive a general relation between the number of activity centers of a city and its population.

## 6.1 FUJITA AND OGAWA

In line with the tradition of economic geography [93], the model of Fujita and Ogawa [91] is based on the concept of agglomeration economies—to explain why economical activities tend to group—and the spatial distribution of wages and rents across the urban space. They consider that cities are constituted of two kinds of actors: the firms, who tend to concentrate to maximise their production, and the households, who try to minimise their rent and commuting cost.

The model is *static*, in the sense that the numbers of firms and individuals are fixed. It is an *equilibrium* model, and considers that the city is the realisation of a general optimum. The original model is also *one-dimensional*, although the hypothesis of one-dimensionality is not fundamental, and only necessary to make the calculations easier. Because we do not try to solve the model, we write equations in the more general two-dimensional case.

### 6.1.1 Households

Fujita and Ogawa assume that there is a fixed number  $N$  of households in the city. The households are considered identical, in the sense that they all have the same utility function and the same budget constraint. The utility function of each household is given by the function  $U = U(Z)$  where  $Z$  is the surplus of money that is left after budgetary constraints (expressed in monetary units); basically, the money one has left at the end of the month, once the rent, bills and petrol (or transportation card) have been paid.

The utility is assumed to be an increasing function of  $Z$  so

$$\frac{\partial U}{\partial Z} > 0 \quad (15)$$

The budget constraint on an household living at  $i$  (of coordinates  $\vec{x}$ ) and working at the firm located at  $j$  (of coordinates  $\vec{y}$ ) is given by the equation

$$Z = W(j) - C_R(i) - C_T(i, j) \quad (16)$$

where  $W(j)$  is the wage earned at  $j$ ,  $C_R(i)$  the total rent paid at  $i$  and  $C_T(i, j)$  the cost of commuting between home and work. This equation is very general, and will be our starting point for the model presented in the next section. The authors of [91] further specify the commuting cost

$$C_T(i, j) = t d_E(i, j) = t |\vec{y} - \vec{x}| \quad (17)$$

where  $t$  represents the commuting cost per unit distance, and  $d_E(i, j) = |\vec{y} - \vec{x}|$  the euclidean distance between home and work. The total rent cost is further written as

$$C_R(i) = R(i) S_h \quad (18)$$

where  $R(i)$  is the rent per unit surface at  $i$ , and  $S_h$  the surface area used by households, which becomes a parameter of the model. The surplus  $Z$  thus finally reads

$$Z = W(j) - R(i) S_h - t d_E(i, j) \quad (19)$$

### 6.1.2 Firms

The second type of agents taken into consideration in the model are the firms. It is assumed that all firms employ the same number of individuals, which amounts to having a fixed number  $M$  of firms (once the

number of households is fixed). The profit earned by a firm located at  $j$  reads, in a general form

$$\Pi = G(j) - C_R(j) - W(j) L_f \quad (20)$$

where  $G(j)$  is the total gain realised by the firm selling its production,  $C_R(j)$  the rent paid by the firm, and  $L_f$  the total number of employees per firm—a parameter of the model.

To take agglomeration economies into account, Fujita and Ogawa define the locational potential  $F$  defined by

$$F(j) = \int_{\mathcal{C}} b(\vec{x}) e^{-\alpha |\vec{y} - \vec{x}|} d\vec{x} \quad (21)$$

where  $b(\vec{x})$  is the density of firms at  $\vec{x}$ . The integral runs over the entire city's spatial extent  $\mathcal{C}$ . One can easily see that the higher the density of firms in a radius of  $1/\alpha$  around a firm, the higher the locational potential is going to be. Balanced by the constraint imposed by the rent, which prevents too many firms from agglomerating at the same location, the locational potential likely is the term responsible for the existence of polycentric solutions in the model. Indeed, the authors further write the total gain  $G$  as a multiple of  $F$ :

$$G(j) = \beta F(j) \quad (22)$$

where  $\beta$  integrates both the productivity of the employees and the effect of the locational potential. The rent, as in the case of households, is written  $C_R(j) = R(j) S_f$  where  $S_f$ , the surface needed by firms, is a parameter of the model. The profit of companies therefore reads

$$\Pi = \beta \int_{\mathcal{C}} b(\vec{x}) e^{-\alpha |\vec{y} - \vec{x}|} d\vec{x} - R(j) S_f - W(j) \quad (23)$$

### 6.1.3 *Equilibrium conditions and results*

Once the budget constraints have been explicated, one needs to define the equilibrium conditions to be able to solve the model. First, the goal of each household is to maximise their utility under the budget constraint. That is, to choose  $Z$ ,  $S$ ,  $\vec{x}$  and  $\vec{y}$  so that  $U(S, Z)$  is maximum.

Here, the maximisation of utility under budget constraints is equivalent to choosing the residential location  $i$  and the job location  $j$  so as to maximise  $Z$ . In other words, the maximisation of utility in this particular situation is equivalent to performing a cost-benefit analysis.

The firms have no utility function, and choose to be at the location  $j$  that maximises their profit.

A further constraint is given by the bid-rent curve, and determines the spatial interaction between households and firms. The authors define two intermediate functions,  $\Psi(\vec{x})$  and  $\Phi(\vec{x})$  which are respectively the bid rent function of households and of firms, defined as

$$\Psi(\vec{x}) = \max_{\vec{x}} \left\{ \frac{1}{S_h} [W(\vec{x}) - Z - t d_E(\vec{x} - \vec{y})] \mid U(Z) = U \right\} \quad (24)$$

$$\Phi(\vec{x}) = \frac{1}{S_f} [\beta F(\vec{y}) - \Pi - W(\vec{y})] \quad (25)$$

$\Psi(\vec{x})$  represents the maximum rent that the households could pay to be located at  $\vec{x}$  while still having a utility value  $U$ .  $\Phi(\vec{y})$  is the maximum rent that firms could pay to be located at  $\vec{y}$ . At equilibrium, it is assumed that whoever's bid rent function has the highest value at  $\vec{x}$  will be located at  $\vec{x}$ .

Taken together, the equilibrium conditions determine the spatial distribution of households and firms, of the wages and land prices.

The results of this model, given its intricacy, are somewhat disappointing. Unsurprisingly, the authors are not able to derive an analytical solution for their model. What they do, however, is deriving the conditions on the parameters for the existence of monocentric and polycentric organisations of activities, using numerical methods.

## 6.2 PROBLEMS WITH THE FUJITA AND OGAWA MODEL

The approach of Fujita & Ogawa fails at giving a satisfactory quantitative account of the polycentric transition of cities. A lot can be said about the details of the model and its assumptions. But we choose to only discuss the issues that we feel are the most important, and that we will try to address in our model.

**IT IS AN EQUILIBRIUM MODEL.** In line with the rest of Urban Economics [93, 92], the authors describe a city as being in an equilibrium characterised by static spatial distributions of households and business firms. However, the equilibrium assumption is unsupported as cities are out-of-equilibrium systems and their dynamics is of particular interest for practical applications [26].

**IT IS TOO COMPLEX.** The model integrates so many interactions and variables that it is difficult to understand the hierarchy of processes governing the evolution of cities: which ones are fundamental and which ones are irrelevant. A model is however only interesting when it



provides a simple structure to understand empirical results, whether it reproduces them, or provides well-understood limiting cases (‘null models’).

IT DOES NOT MAKE ANY PREDICTION. Worse, due to its complexity, the model is unsolvable, and does not make any prediction. At best it shows that polycentric configurations are *possible*. Yet, there are possibly different models that would admit polycentric activity profiles as a solution. The constraint is not strong enough, the model is unsupported by data.

We also note that the model does not take congestion into account in the commuting cost (which is only a function of the distance). However, as we saw in Chapter 5, it is mentioned in the economics literature as being a possible cause of the polycentric transition of cities [157].

### 6.3 MODELING MOBILITY PATTERNS

In this section, we start from the model by Fujita & Ogawa to propose a dynamical model of city growth. Following recent interdisciplinary efforts to construct a quantitative description of cities and their evolution [144, 233, 148, 39, 26], we deliberately omit certain details and focus instead on basic processes. We thereby aim at building a minimal model which captures the complexity of the system and is able to account for – qualitative as well as quantitative – stylized facts.

The model we propose is by essence dynamical and describes the evolution of cities’ organisation as their population increases. We focus on car congestion – mainly due to journey-to-work commutes – and its effect on the job location choice for individuals.

#### 6.3.1 *Decoupling the choice of household location and job*

The time scales involved in the evolution of cities are usually such that the employment turnover rate is larger than the relocation rate of households. On a short time scale, we can thus focus on the process of job-seeking alone, leaving aside the problem of the choice of residence. In other words, we assume the coupling between both processes to be negligible: we assume that each inhabitant newly added to the city has a random residence location and we concentrate on understanding how such an inhabitant chooses its job location.

As a result of this assumption, a worker living at  $i$  will choose to work at the center  $j$  such that the quantity

$$Z_{ij} = W(j) - C_T(i, j) \quad (26)$$

is maximum. Doing so, we give up any hope to describe the spatial structure of the rent distribution, or the alledged scaling between rent prices and population size in cities [38].

### 6.3.2 *Decoupling the behaviour of firms and individuals*

Another difficulty with the Fujita-Ogawa model is the strong coupling between the behaviour of firms and individuals. The empirical literature on the behaviour of firms points to a tendency of similar industries to cluster geographically [74, 146], and a higher profit of industries located in Urban environments [158]. Although theoretical attempts at explaining these behaviours have been proposed [75], the models are yet to be developed in an out-of-equilibrium framework.

Here, we decide to simplify the problem by assuming that firms indeed cluster into specific locations, that we call activity centers. Each worker can then choose among a pool of  $N_c$  potential activity centers (whose locations are randomly distributed across the city). The active subcenters are then defined as the subset of potential centers which have a non zero incoming number of individuals. We thus assume that the existence of activity centers is defined by the willingness of workers to work in the possible locations.

Let us now discuss the form of the wage  $W(j)$  and the commuting cost  $C_T(i, j)$  that are present in equation 26.

### 6.3.3 *Determining the wage*

The problem of determining the (spatial) variations of the average wage  $W(j)$  at location  $j$  is very reminiscent of some problems encountered in fundamental physics. Indeed, the wage depends on many different factors, ranging from the type of company, the education level of the inhabitant, the level of agglomeration, etc., and in this respect is not too different from quantities that can be measured in a large atom made of a large number of interacting particles. In this situation, physicists figured that although it is possible to write down the corresponding equations, not only is it impossible to solve them, but also not really useful. In fact they found out that a statistical description of these systems, relying on random matrices could lead to predictions which agree with experimental results [77].

We wish to import in spatial economics this idea of replacing a complex quantity such as wages – which depends on so many factors and interactions – by a random one. The problem is not so much that we cannot write down the equations that determine the wage that an individual could get in a given company. Even if we could (and we can't),

the sheer number of people living in an urban area would prevent us from solving these equations. And even if we could solve them, the resulting information would be too overwhelming to really allow us to *understand* the behaviour of the system as a whole. We thus need an *effective* description of the phenomenon.

We account for the interaction between activity centers and people by taking the wage in location  $j$  as proportional to a random variable  $\eta_j \in [0, 1]$  such that  $W(j) = s \eta_j$  where  $s$  defines the maximum attainable average wage in the considered city.

We are aware that wages are not determined endogenously but are instead the result of thousands, millions of interactions between firms and individuals. In the same way that Dyson did not mean that the interactions between electrons in large atoms *are* random, our assumptions does not mean that wages are *really* randomly determined. What we mean, however, is that in the case of systems containing a large number of individuals, one may do *as if they were randomly determined*. Although we thereby abandon the possibility to describe the dynamics of the wages and their spatial distribution, the resulting model is analytically solvable and makes quantitative predictions.

#### 6.3.4 *Commuting cost and congestion*

We choose the transportation cost  $C_T(i, j)$  proportional to the commuting time between  $i$  and  $j$ . In a typical situation where passenger transportation is dominated by personal vehicles, this commuting time not only depends on the distance between  $i$  and  $j$ , but also on the traffic between the two places, the vehicle capacity of the underlying network and its resilience to congestion. The Bureau of Public Road formula [47] proposes a simple form taking all these ingredients into account. In our framework, it leads to the following expression for the commuting costs

$$C_T(i, j) = t d_{ij} \left[ 1 + \left( \frac{T_{ij}}{c} \right)^\mu \right] \quad (27)$$

where  $T_{ij}$  the traffic per unit of time between  $i$  and  $j$  and  $c$  is the typical capacity of a road (taken constant here). The quantity  $\mu$  is a parameter quantifying the resilience of the transportation network to congestion. We further simplify the problem by assuming than the traffic  $T_{ij}$  is only a function of the subcenter  $j$  and therefore write  $T_{ij} = T(j)$  the total traffic incoming in subcenter  $j$ .

#### 6.3.5 *Summary*

In summary, our model is defined as follows. At each time step, we add a new individual  $i$  located at random in the city, who will choose to

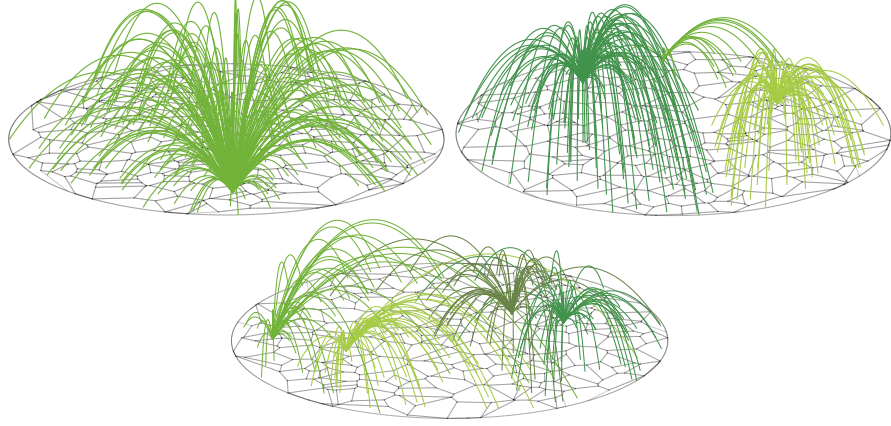


Figure 12: **The different regimes.** The monocentric (top left), distance-driven polycentric (top right) and attractivity-driven polycentric (bottom) regimes as produced by our model. Each link represents a commuting journey to an activity center.

work in the activity area  $j$  (among  $N_c$  possibilities located at random) such that the following quantity

$$Z_{ij} = \eta_j - \frac{d_{ij}}{\ell} \left[ 1 + \left( \frac{T(j)}{c} \right)^\mu \right] \quad (28)$$

is maximum (we omitted irrelevant multiplicative factors). The quantity  $\ell = s/t$  is interpreted as the maximum effective commuting distance that people can financially withstand. Interestingly, the presence of commuting costs entails the existence of a second length scale  $\ell$  in the system (the first one being the typical size  $L$  of the city).

#### 6.4 MONOCENTRIC TO POLYCENTRIC TRANSITION

Depending on the relative importance of wages, distance and congestion, the model predicts the existence of three different regimes: the monocentric regime (Top left Figure 12), the distance-driven polycentric (Top right Figure 12) regime and the attractivity-driven polycentric (Bottom Figure 12) regime.

The existence of a monocentric regime depends on how  $\ell$  – the maximum commuting distance that people can afford – compares to the size of the city  $L$ . Indeed, people located at a distance  $d > \ell$  from the most attractive center will not be able to afford commuting to this center, and will, according to our model, choose to commute to a closer center. As a result, a monocentric regime is only sustainable as long as people’s residence is drawn close to the most attractive center. Thus, in the limit where  $\ell \ll L$ , the attractiveness of a center becomes irrelevant, and a monocentric regime cannot exist. In this case, we end up in the situa-

tion shown on the top-right of Figure 12.

From now on, we will assume that  $\ell$  is large enough so that a monocentric state exists for small values of the population. In this regime, the value of  $\eta$  prevails and the monocentric state evolves to an attractivity-driven polycentric structure as the population increases. Starting from a small city with a monocentric organisation, the traffic is negligible and

$$Z_{ij} \approx \eta_j$$

which implies that all individuals are going to choose the most attractive center, with the largest value of  $\eta_j$ , say  $\eta_1$ . When the number  $P$  of individuals increases, the traffic will also increase and some initially less attractive centers (with a smaller values of  $\eta$ ) might become more attractive, leading to the appearance of a new subcenter. More specifically, a new subcenter  $j$  will appear when for an individual  $i$ , we have

$$Z_{ij} > Z_{i1}$$

Because we assumed we originally were in a monocentric state, the traffic at this point is such that  $T(1) = P$  and  $T(j) = 0$  which leads to the equation

$$\eta_j - \frac{d_{ij}}{\ell} > \eta_1 - \frac{d_{i1}}{\ell} \left[ 1 + \left( \frac{P}{c} \right)^\mu \right] \quad (29)$$

We assume that there are no spatial correlations in the subcenter distribution, so that we can make the approximation  $d_{ij} \sim d_{i1} \sim L$ . The new subcenter will thus be such that  $\eta_1 - \eta_j$  is minimum. It will thus be the potential subcenter with the second largest value denoted by  $\eta_j = \eta_2$ .

According to order statistics, we have on average for a uniform distribution

$$\overline{\eta_1 - \eta_2} \simeq 1/N_c$$

hence a critical value for the population

$$\boxed{P^* = c \left( \frac{\ell}{LN_c} \right)^{1/\mu}} \quad (30)$$

Whatever the system considered, there will *always* be a critical value of the population above which the city becomes polycentric. The monocentric regime is therefore fundamentally unstable with regards to population increase, which is in agreement with the fact that no major city

in the world exhibits a monocentric structure. We note that the smaller the value of  $\mu$  (or larger the value of the capacity  $c$ ), the larger the critical population value  $P^*$  which means that cities with a good road system capable of absorbing large traffic should display a monocentric structure for a longer period of time.

### 6.5 NUMBER OF CENTERS

We have so far established that, because of increased levels of congestion as the population grows, all cities will eventually adopt a polycentric structure. Although appealing and in agreement with common observations, the prediction given by Eq. 30 is impossible to test with the currently available data. Therefore, we would like to obtain a prediction for the variation of the number of subcenters with population.

We compute the value of the population at which the  $k^{th}$  center appears. Still in the attractivity-driven regime, we assume that so far  $k-1$  centers have emerged with

$$\eta_1 \geq \eta_2 \geq \dots \geq \eta_{k-1}$$

with a number of commuters  $T(1), T(2), \dots, T(k-1)$ , respectively. The next worker  $i$  will choose the center  $k$  if

$$Z_{ik} > \max_{j \in [1, k-1]} Z_{ij} \quad (31)$$

which reads

$$\eta_k - \frac{d_{ik}}{\ell} > \max_{j \in [1, k-1]} \left\{ \eta_j - \frac{d_{ij}}{\ell} \left[ 1 + \left( \frac{T(j)}{c} \right)^\mu \right] \right\} \quad (32)$$

According to simulations of the model, we know that the distribution of traffic  $T(j)$  is narrow [2], and we can assume that all the centers have roughly the same number of commuters  $T(j) \sim P/(k-1)$ . As above we also assume that there are no spatial correlations in the position of employment centers so that  $d_{ij} \sim d_{ik} \sim L$ . We can now write the previous expression as

$$\frac{L}{\ell} \left( \frac{P}{(k-1)c} \right)^\mu > \max_{j \in [1, k-1]} (\eta_j) - \eta_k \quad (33)$$

Following our definitions,  $\max_{j \in [1, k-1]} (\eta_j) = \eta_1$ . According to order statistics, if the  $\eta_j$  are uniformly distributed, we have on average

$$\overline{\eta_1 - \eta_k} = (k-1)/(N_c + 1)$$

It follows from these assumptions that (1) the  $k^{th}$  center to appear is the  $k^{th}$  most attractive one (2) the average value of the population  $\bar{P}_k$  at which the  $k^{th}$  center appears is given by:

$$\bar{P}_k = P^* (k-1)^{\frac{\mu+1}{\mu}} \quad (34)$$

Conversely, the number  $k$  of subcenters scales sublinearly with population size as

$$k \sim \left( \frac{P}{P^*} \right)^{\frac{\mu}{\mu+1}} \quad (35)$$

For positive values of  $\mu$ , we have  $\frac{\mu}{\mu+1} < 1$ . we can thus conclude that the number of activity subcenters in urban areas scales sublinearly with their population where the prefactor and the exponent depend on the properties of the transportation network of the city under consideration. This prediction is in agreement with the scalings obtained for Spanish and American cities in Chapter 5.

## 6.6 CONCLUSION

### 6.6.1 A predictive model

The model we just presented, although not perfect, exhibits many of the desirable features of a model we listed in the introduction. First, it goes beyond the standard models in urban economics by going beyond the explanation of simple, qualitative, stylized facts. As we saw earlier, one major problem with the model of Fujita and Ogawa is the absence of quantitative prediction. Instead of providing a prediction that can be further confirmed or refuted by empirical observation, the authors merely test the existence of polycentric solutions in the framework of their model. The link with reality is however very loose, in the sense that there is a big intellectual leap between the actual prediction of the model and reality. Even though the model proposed here is very simple, it is not difficult to link it to reality. Once the notion of activity centers is defined empirically, it is not difficult to count the number of centers and look at the dependence of this number on the population size of cities. The model can then be confirmed, or refuted. Furthermore, as we will see in the following section, the model serves as a basis to the understanding of some of the scaling relationships in cities, linking the model even more strongly to empirical reality.

### 6.6.2 Understanding the polycentric transition

Second, the model allows us to *understand* why the polycentric transition occurs. Taking a step back on the assumptions that lead to the

prediction of Eq. 35, one can see that the transition in our model is triggered by the congestion term in Eq. 28. The positions of households and firms are indeed taken as random, the wages are also taken at random. Therefore, we can conclude that our model explains the polycentric transition of cities through the increasing congestion around employment centers as the population increases. More mechanisms are probably involved, but the model shows that congestion alone is enough to lead to a polycentric situation.

If we assume that agglomeration economies can explain the existence of centers in the first place, the model provides evidence that this centripetal force is balanced by the centrifugal effect of congestion that tears cities apart. Arguably, the non trivial spatial patterns observed in large cities can be understood as a result of the interplay between these competing processes.

The model we propose trades off exhaustivity and complexity for simplicity and explanatory power. Although some of the hypotheses we made are debatable, it is striking that we manage to make a prediction on the scaling of the number of centers with population size. On the other hand, unlike simplistic model, our model's ontology is hard-wired into the reality we experience. For this reason, its assumptions can be discussed, possibly changed. The model can be improved upon in many different ways.



*Our progress is narrow;  
it takes a vast world unchallenged and for granted.*

— J. Robert Oppenheimer [173]

As we stated in the introduction, all models are fundamentally wrong – at least incomplete. Although it is able to reproduce key empirical regularities, the model presented in Chapter 6 is no exception to this rule. In the following chapter, we will enumerate some of its weaknesses, and propose possible ways in which it could be extended.

Besides, because they are trying to make sense of a complex reality with a limited number of tools, empirical analyses are not exempt of limitations either. Before closing this chapter, we question the validity of the distribution-based methods used to identify subcenters, and challenge the notion of polycentricity itself.

## 7.1 QUESTIONING AND EXTENDING THE MODEL

### 7.1.1 *What the model does not say*

The model makes many simplifying assumptions that make it analytically tractable, but hide some interesting aspects of intra-urban dynamics. We do not pretend to explain the complexity of urban dynamics in its entirety, but rather some of its aspects.

A first feature, hidden in the assumptions of the model, is that we do not explain the concentration of activities in particular areas of the cities. Rather, we take the existence of centers for granted, and do not bother with the behaviour of firms. Of course, this is a topic worthy of investigation, and should be studied in more depth in order to have a comprehensive understanding of the mechanisms that shape cities.

A second limitation lies in the fact that we ignore the process of residence choice, and attribute households' location at random in the city. We therefore set aside the problem of competition for space between households, and a theoretical description of the spatial distribution of housing prices (see [97] for a model that explores this aspect).

Another limitation lies in the description of congestion. In a worry to simplify the problem, we chose to adopt a macro-scale description of traffic congestion, given by Eq. 27. The sensitivity of the road network to congestion is taken into account through the exponent  $\mu$  and the capacity  $C$ , which are assumed to be the same across the entire city.

In order to derive and compute these parameters, one would need to understand how local patterns of congestion lead to macroscopic behaviours at the city scale. This is, of course, a difficult enterprise: local particularities of the layout may have dramatic consequences on the fluidity of traffic, and congestions do propagate through the network so that access to a given center can have an effect on the travel to another center [139].

### 7.1.2 Possible avenues

Even without considering the difficult problem of modeling the behaviour of the firms, and the way it is coupled to that of individuals, the model could be improved in several ways. One first possible extension is to take the presence of public transportation into account. Indeed, the model only considers individual vehicles, prone to congestion, as a transportation mean. However, the largest cities in the world are all served by metro systems [195], and the share of transports other than personal vehicles can attain 42% in cities like New-York. It is therefore far from being negligible, and should be taken into account in the model. In its defense however, cars remain the dominant mode of transportation in the U.S., as shown of Figure 13. The use of alternative modes of transportation is only notable in New-York, which is already a poly-centric city.

Number from the  
2013 American  
Community  
Survey

Another possible (but non-trivial) extension to the model is linked to the second limitation stated above. Adding an income structure into the model (and rules concerning the interaction of individuals) could allow us to explore the spatial patterns of segregation, and see whether they can be understood from basic economical choices alone [97]. We considered this avenue during this thesis, and realised there was very little of the empirical knowledge on segregation could be used to test a model. This led us to working on the material presented in Part iv of this thesis.

## 7.2 SHADOWS IN THE EMPIRICAL PICTURE

### 7.2.1 Identifying and counting centers

Although non-parametric methods are an improvement over the previous parametric methods, we are yet to understand the exact meaning of the obtained centers.

In particular, a problem that remains with non-parametric methods is that, no matter the distribution of employment, population, etc. into the areal units, the method will output a number. For instance, let us consider the extreme case of a city where employment is uniformly distributed in space, so that the employment density is uniform. In this

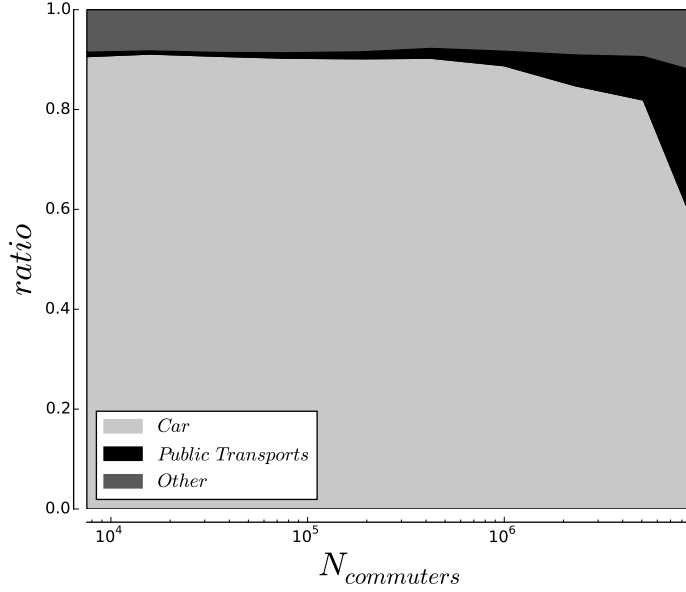


Figure 13: **Mode share in the U.S..** Importance of different transportation modes in U.S. Metropolitan Statistical Areas, as a function of the number of commuters. Although the proportion of individuals using public transportation or other modes (walking, cycling, working at home) increases with population size, cars stay the dominant mode of transportation everywhere. Data are from the 2013 American Community Survey.

situation, the LouBar method would tell us that the number of centers is equal to the number of areal units. Yet, can we really talk about centers in this case? Most would (rightfully) object. But on what ground?

The difficulty resides in that we do not know what we mean exactly when we talk about centers: do they reflect an objective reality, or are they a mere artifact of the way our brains process information? Can they be quantitatively defined, based on their desired properties or are they merely ‘unusual’ fluctuations in the distribution of activities? In the latter case, parametric methods will do just fine. In the former case means we need to understand what we talk about when we talk about centers. It is somewhat ironic that, more than 15 years after the publication of McDonald’s seminal paper [155], we are still pondering over the question he originally asked.

A further shortcoming of the most recent (distribution-based) methods is that they do not consider the spatial arrangement of the areal units involved. This can be problematic, especially when the method identifies as centers areal units that are contiguous.

We show an example of such a situation on Figure 14. We use the LouBar method [141] to extract the employment hotspots in the Boston, MA MSA using data from the 2000 Census. As one can see, several of

the identified hotspots are contiguous. Should we still count them as separate hotspots? Or should we consider that all contiguous hotspots are part of a larger hotspot that encompasses them all?

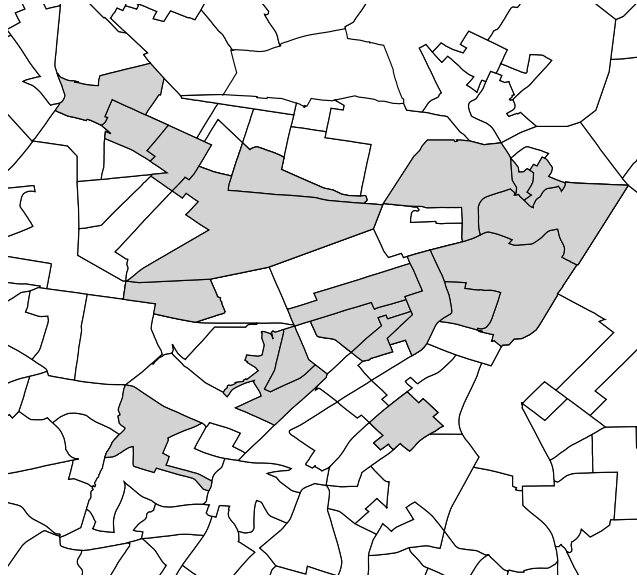


Figure 14: **Subcenters and contiguity.** The census tracts of downtown Boston, MA in the U.S.. In light grey, the census tracts that are identified as employment hotspots by the LouBar method. Although the method designates all light grey tracts as different hotspots, many of them are contiguous. We can wonder whether such contiguous hotspots are, in fact, part of a larger hotspot that would include all of them. This plot was generated using the 2000 Census tract-to-tract commuting flows and the 2000 Census tracts geometry.

The results of the methods provided in the introduction should not be thrown away altogether, though. The number of centers they provide probably does not reflect the ‘real’ number of centers (if there is such a thing) in a particular city. But, assuming that different cities exhibit similar structures, they should still provide values that are coherent across different urban areas, and are thus useful for *comparison* purposes.

## 7.2.2 *Beyond polycentricity?*

### 7.2.2.1 *The dispersed city*

As we saw in Chapter 5, the concept of the monocentric city was progressively replaced with the more elaborate polycentric hypothesis. It is, however, not the end of the story. Gordon and Richardson, in a

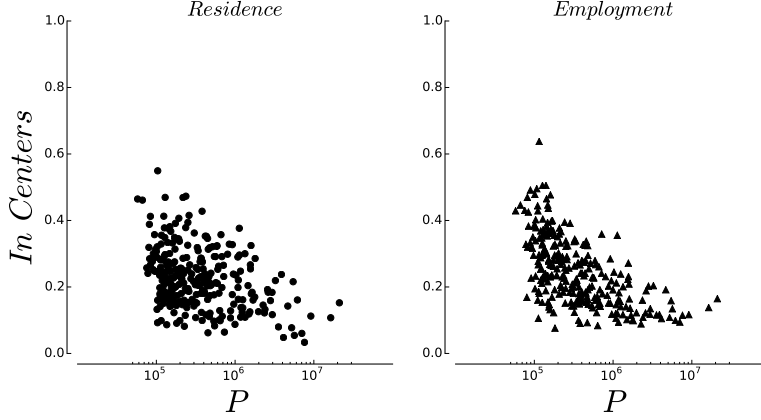


Figure 15: **Concentration in subcenters.** (Left) Ratio of the total residential population in U.S. MSAs that lives in the centers identified by the LouBar method. (Right) Ratio of the total number of employees in U.S. MSAs that work in the centers identified by the LouBar method. Overall, cities are very dispersed, with only a few cities having more than 50% of their workforce or residential population living in centers, confirming the results of Gordon and Richardson [103]. Data are from the 2000 U.S. Census.

provocative article [103], argue that cities are dispersed more than they are polycentric. Indeed, studying the employment density in Los Angeles, they found that the centers they identified only contained 17% of the total employment. Hardly a polycentric situation!

Of course, we can (and should) wonder whether Gordon and Richardson’s results are an artefact of the choice of their case study –Los Angeles, famous for its sprawl– or the particular method they used to compute the number of centers. We thus plot on Figure 15 the ratio of the total number of individuals that is contained in the centers defined by the LouBar method. The results are striking: only a few, small metropolitan area reach the mark where 50% of individuals (employees or residents belong) to a designed center. Worse, cities seem to be on average more dispersed as they are bigger.

The lesson that should be learned from the article by Gordon and Richardson is that the notion of polycentricity is *also an hypothesis* on the spatial structure of densities. While it is arguably more involved than the monocentric hypothesis, it does indeed implicitly impose some structure onto the data. The process itself of counting centers implies that these centers exist, that there is an element of reality attached to what we call centers. A quick look on the 3D plot shown on Figure 6 should convince the reader that the world is not as simple as the way we picture it. For instance, while employment densities indeed exhibit strong peaks that are easily distinguishable (although that is arguable for Houston), the same cannot be said for population densities.

The point is not that the monocentric or the polycentric model are wrong altogether. The problem lies in the lack of appropriate tools to describe a density spatial profile, in the fact that there is no ‘one size fits all’, unbiased method of analysis. Indeed, the exploratory tools presented above try to fit a certain model of the city to the actual data, be it monocentric or polycentric. The methods developed to identify centers count the centers *provided* there are centers. We definitely need more elaborate methods that are also able to tell us *whether* there are centers. Or that go beyond the notion of center.

#### 7.2.2.2 *Quantifying Urban form*

This problem is in fact very general, and pertains to the field of spatial analysis (including spatial statistics). Finding centers indeed amounts to finding the proper way to describe a density profile at a meso-scale level and to devising proper methods to detect the salient feature of this spatial pattern. The collection of tools and methods to describe the structure of density patterns in cities constitutes the sub-field of urban form [222, 205, 130, 131] and reaches far beyond the determination of subcenters.

Finally, we have focused in this part on the *morphological* aspect of urban form, as most of the preceding studies. We acknowledge however the existence of a *functional* aspect (see [33]), which takes the attraction range of employment subcenters into account, in addition to the raw number of employees. Mixing employment densities and the property of the flows to the center may indeed lead to a better understanding of what a center really is.

### 7.3 SUMMARY

In this part, we have presented an historical overview of the monocentric hypothesis for the structure of cities, and how the view has progressively shifted towards the picture of a more distributed, polycentric organisation. Starting with indirect evidence for a polycentric picture, several methods were then naturally proposed to directly measure the number of centers, from the first parametric methods to the more recent non-parametric methods. Observing evidence for an increased polycentricity with population size, we then wondered what were the possible explanations for this phenomenon. We proposed an out-of-equilibrium model of city growth that predicts the necessary emergence of secondary centers as populations grows, and a sublinear increase of the number of subcenters with population—both verified on empirical data, across different countries, for several city definitions.

In the next part, we will continue our journey with another, seemingly unrelated topic: scaling relationships. We will start with a historical perspective on scaling, showing that scaling relationships did in fact precede Quantitative Geography, and we will provide a non-exhaustive

review of the empirical results. We will then be ready to show how, using the model exposed in the previous chapter, we can understand the value of the scaling exponents related to individual mobility. We will then conclude on a reflection of what scaling relationships can and do tell us about cities, and highlight their shortcomings.





## Part III

### SCALING

The past decade has witnessed a renewed interest for the scaling of some of cities' characteristics with population size – first discovered more than 60 years ago.

The contribution of this part is threefold. First, we review the existing literature on allometric scalings, sorting the measured exponents by theme. We then propose a model to explain the scaling exponent of several indicators related to mobility in cities, and discuss the theoretical and practical consequences of these exponents. Finally, we present some of the challenges posed by scaling relationships: their interpretation, and the issues they reveal about the definition of cities.



## INTRODUCTION

*The allometric law promises to become  
an integral part of geography theory.*

— David Harvey (1969) [108]

## 8.1 PROBING CITIES WITH SCALING LAWS

8.1.1 *Scaling laws*

As discussed in the introduction of this thesis (Chapter 1), cities are paradigmatic examples of complex systems. As systems, they can be of thought of as ‘black boxes’ with inputs (people, goods, money, information, etc.), a structure (roads, buildings, electric cables, etc.) and outputs (Patents, CO<sub>2</sub> emissions, etc.). A simple way to explore the behaviour of such a system is to look at the way it behaves when we change its size. That is, how its structure and its outputs change when the inputs are altered. Formally speaking, we try to find the function  $f$  such that the quantity  $Y$  – a measure of the output or the structure – varies as

$$Y = f(S) \tag{36}$$

where  $S$  is the size of the system.

What is to be considered as the size of the city? The spatial footprint, the total volume occupied by its building? The answer adopted by many before this thesis [214, 39], is the total number of inhabitants. The real reason is probably pragmatic: “it works”. Although, in retrospect, the choice of population makes complete sense.

Cities are indeed more than roads and buildings: cities are the people who inhabit them. People are responsible for the changes in wealth, employment, number of patents. People need new roads, and it is people who build them. People need electricity, and again it is people who run electric cables between buildings. Inhabitants of a city, through their actions and interactions, are responsible for the collective mechanisms that act on the city as a whole. In a sense, behind the use of the population  $P$  to measure the size of a city as a system hides the idea that cities are, first and foremost, the people that inhabit them.

As a matter of fact, when we try to plot quantities as a function of the population size  $P$  of cities, we obtain *allometric scaling relationships*.

That is, a power-law relationship between various quantities  $Y$  and the population size  $P$  of cities *in a given system of cities*

$$Y = Y_0 P^\beta \quad (37)$$

where the exponent  $\beta$  can be different from 1. This type of scaling relation, used extensively in Biology [219] and in Physics [21], is a signature of the various processes governing the phenomenon under study, especially when the exponent  $\beta$  is different from what would be naively expected. Three qualitatively different regimes are usually distinguished for the exponent  $\beta$  [39]

**SUPERLINEAR** when  $\beta > 1$ . In this situation, the  $Y$  per capita increases with population size. This is associated with the notion of increasing returns with scale in economics.

**LINEAR** when  $\beta = 1$ . In this situation, the  $Y$  per capita is constant. This behaviour is characteristic of an extensive system, when the whole is equal to the sum of its parts.

**SUBLINEAR** when  $\beta < 1$ . In this situation, the  $Y$  per capita decreases with population size. When  $Y$  is the cost in infrastructure, this is characteristic of economies of scale.

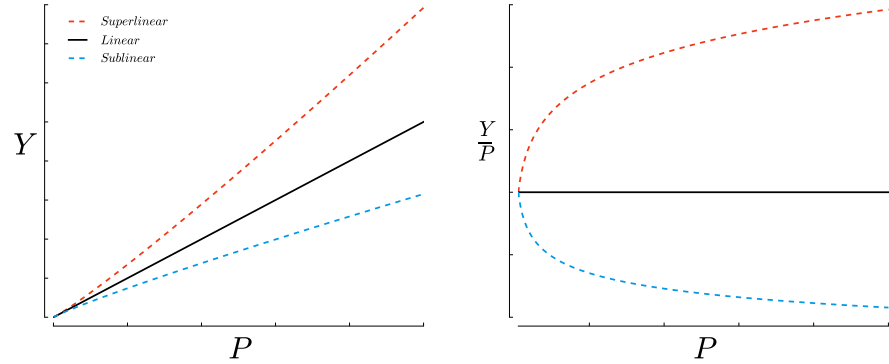


Figure 16: **Sublinear, Linear and Superlinear scaling.** (Left) Example of a linear (black), sublinear (blue) and superlinear (red) behaviour. (Right) Evolution of the correspondent per-capita quantities with population. A superlinear behaviour means that per-capita quantities increase with population size, while a sublinear behaviour means per-capita quantities decrease with city size.

We note that the scaling exponent  $\beta$  is also directly related to the *elasticity* defined in Economics. Indeed, the cities' population elasticity of the quantity  $Y$  is defined as

$$\beta = \frac{dY/Y}{dP/P} \quad (38)$$

### 8.1.2 *Underlying assumptions*

Several assumptions, although rarely mentioned, hide behind every exhibited scaling law. The first one, is that we are able to unambiguously delineate cities as systems. While this is trivial in the case of animals (it is fairly easy for us to isolate an elephant, or a cat from its environment before measuring its mass and its metabolic rate), it is a much more difficult task in the case of cities. Indeed, cities do not have fixed boundaries, and their geographical limits evolve with time. They are also open system: people are born and die, change residence and companies do the same.

Traditionally, people have relied on the definition given by statistical agencies of the respective countries they were studying – and we will do the same in the next chapter. We will however see, in the chapter concluding this part, that the problem of delineating cities is a sensible issue and affects greatly scaling analyses.

A second issue, rarely – if ever – mentioned in the literature, is the necessity to define the set of cities to study. Scaling laws are essentially cross-sectional relationships, where we measure the quantity  $Y$  on a set of cities with different populations. But how is the set determined? For instance, would it make sense to mix French cities, Ukrainian, Canadian and Korean, etc cities and plot, say, their total GDP as a function of the population? Would we then observe a neat scaling relationship?

Intuitively, this is very unlikely to happen, as different countries have overall different levels of wealth, and this should be reflected in the wealth of their cities. Therefore, plotting cities from different countries together is likely to introduce important deviations to the pure scaling relations which are not due to the fact that cities in different countries do not follow the same processes, but rather because of systemic differences at the country level. As a matter of fact, most studies limit themselves to a single country. But one should bear in mind that this choice is arbitrary. And the problem of choosing the appropriate set from which to pick the cities is linked to the more general problem of defining systems of cities.

### 8.1.3 *An increasing importance*

This chapter's epigraph, from Harvey's 1969 *Explanation in Geography*, is somewhat prophetic. Allometric scaling relationships only concern 1 page out of the 500 pages that the book contains, a reflection of the very few empirical results that were available at the time. Looking at the extent of the literature on scaling relationships almost 50 years after Harvey wrote this sentence, it is difficult to deny the accuracy of this prophecy. Thanks to the wider availability of data through statistical agencies, but also the availability of 'new data' (such as mobile phone

data), empirical measurements of scaling laws have multiplied, and now concern quantities as diverse as the total surface area, the number of new patents, the quantity of CO<sub>2</sub> emitted, the number of phone contacts of individuals, etc. The discovery of allometric scaling in cities is not recent [214], but it has undoubtedly caused a stir in the literature about urban systems over the last decade [39, 183, 38, 4, 5, 16].

In the next section, we will present a non exhaustive historical review of the empirical results on scaling relationships. This will lay the ground for our contribution to the debate: a theoretical interpretation of the scalings related to the mobility of people, and an estimate for the scaling exponent of the surface area.

## 8.2 A BRIEF HISTORY OF ALLOMETRIC SCALING AND CITIES

Rather than an exposition that is linear in time, we deliberately choose to classify the proposed studies according to the type of quantity. That way, we emphasize the variety of variables that have been studied. Incidentally, this order also reveals the different waves of interest scaling relationships have sparked off in the past 6 decades, and hints at some issues related to scaling laws.

### 8.2.1 *Surface area*

The spatial footprint of cities, as can be observed on satellite picture or on maps, is one of the properties that is easiest to measure. It is therefore not surprising that the first occurrence of scaling relationships in cities was the scaling of the surface area of cities with their population. In 1947, using data about administrative cities obtained from the 1940 US Census, John Stewart shows

*Incidentally, the  
author of the  
study, John  
Stewart, was a  
physicist.*

$$A = \frac{p^{3/4}}{350} \quad (39)$$

The next occurrence of this scaling can be found 9 years later in a study by the same author [216], using UK census data. It isn't long until the result percolates in Geography with Boyce in 1963 [46]. In 1965, Nordbeck's paper [166] also studies the scaling of surface area with population, and, for the first time, explicitly refers to allometry in biology. Later, Tobler [220] uses some of the first available satellite images to provide the first confirmation using satellite pictures. Satellite pictures were also used more recently by Guérois in [105] (Table 1).

When applied to morphological definitions of cities, all studies (see [29]) give an exponent that varies in the range [0.70, 0.90]. However, different results are obtained for functional definitions of cities [29], or when the

set of studied cities span several systems of cities [94]. Thus, despite being the oldest and most trusted scaling relationship in the literature, the relation between the surface area and population size of cities exhibits some of the issues we will discuss in Chapter 10.

Exponent	City definition	Year	Study
0.75	Administrative (US)	1940	Stewart [214]
0.75	Administrative (UK)	1951	Stewart & Warntz [216]
0.86	Morphological (US)	1950	Boyce [46]
0.88	Administrative (US)	1950	Nordbeck [166]
0.88	Built-up (US)	1969	Tobler [220]
0.86	Built-up (Europe)	1990	Guérouis [105]
0.73	Administrative (Europe)	1990	Guérouis [105]
0.78	Morphological (US)	2010	Louf & Barthelemy [4]
1.48	<i>Functional</i> (US)	2005	Batty & Ferguson [29]

Table 1: **Scaling of the surface area.** Scaling exponents for the surface area of cities found in the literature. The scaling for administrative cities, built-up areas or cities defined according to a morphological criterion are consistent with one another – at least qualitatively. The exponent for cities with a functional definition is however qualitatively different.

### 8.2.2 Economic diversity and employment

#### 8.2.2.1 Employment diversity

The economic diversity has been of interest to researchers very early on. In 1949, Zipf in *Human behavior and the principle of least effort* [234] plots the number of service-business establishments, manufactures and retail stores per city as a function of population (in log-log scale) using data from the 1940 US Census. He finds a linear relationship with population for the three types of establishments, which agreed at the time with his model. He also plots the scaling of the diversity, defined as the number of different kinds of enterprises present in the city being studied.

In his 1967 *Geography of market centers and retail distribution* [35] Berry, hoping to demonstrate the hierarchical organisation of central places, plots this time the population of cities as a function of the number of retail and service businesses observed. Strangely enough, the data imply

$$D \propto P^\beta \quad (40)$$

with  $\beta > 1$ , in contradiction with later results. Indeed, Bettencourt et al. [41] showed that the professional diversity  $D$ , measured as the number of professions of different kind in the city considered, could be fitted by the following function

$$D(N_e) = d_0 \frac{\left(\frac{N_e}{N_0}\right)^\gamma}{1 + \left(\frac{N_e}{N_0}\right)^\gamma} \quad (41)$$

where  $d_0$  is the size of the classification used in the data,  $N_0$  is the typical saturation size, and  $\gamma < 1$  is an exponent expressing the extent to which new activities ‘appear’ as the total employment increases. Far from the saturation regime, when  $N_e \ll N_0$  (the classification is sufficiently fine-grained), we have

$$D(N_e) \sim A N_e^\gamma \quad (42)$$

#### 8.2.2.2 *Employment in different activities*

More recently, Pumain and coauthors [187], extending the work done by Paulus in his PhD thesis [176], showed that the employment  $E_a$  in different activities  $a$  scaled as

$$E_a \propto P^\beta \quad (43)$$

with different exponents  $\beta$  for the different activities (Table 2). They observed, for the year 1999 in France, that the exponents could be classified in three categories

- $\beta > 1$  for innovative sectors: research and developement, consultancy.
- $\beta = 1$  for common sectors: hotels, health and social services, education.
- $\beta < 1$  for ‘mature’ sectors such as the food industry

This result was confirmed recently by Youn et al. [231] – although they do not refer to this previous work – who showed that the same behaviour was observed for the number of business of a given type.

A particularly interesting result by Pumain et al. [187] is the evolution of the different exponents with time, where we can see a clear increase of the exponents for research and developement, and a clear decrease of the exponents related to manufactures of different kinds. We will come back to the interpretation of this phenomenon in Chapter 10.



Exponent	City Definition	Economic sector
1.67	Functional (France)	Research and development
1	Functional (France)	Hotels and restaurants
0.85	Functional (France)	Manufacture of food products

Table 2: **Scaling of employment in different economic sectors.** The scaling behaviour of the number of employees in a given economic sector depends on the nature of the economic sector. We give an example for each of the ‘innovative’ (superlinear), ‘common’ (linear) and ‘mature’ (sublinear) categories defined by Pumain et al. [187]. The exponents were obtained from [187] and concern French 1999 ‘Aires urbaines’.

### 8.2.3 *Wealth*

The notion of increasing returns with the size of the agglomeration is often discussed in economics, although empirical proofs are hard to find. The superlinear scaling of the GDP of american cities as a function of their population may be the most striking example of such increasing returns [39]. In the same article, Bettencourt et al. showed that the number of patents (used as a proxy for creativity), and wages also scaled superlinearly with population size in the US (see Table 3).

Because larger cities create proportionally more wealth than smaller cities, we can wonder whether this supplement of wealth allows to sustain proportionally more jobs. The answer, as shown in [41] for american cities, is negative: the total employment of a city is on average proportional to its population.

Quantity	Exponent	City Definition	Study
GDP	1.13	Functional (US)	Bettencourt [38]
New patents	1.27	Functional (US)	Bettencourt et al. [39]
Total wages	1.12	Functional (US)	Bettencourt et al. [39]
Employment	1.01	Functional (US)	Bettencourt et al. [39]

Table 3: **Economic vitality.** The scaling of quantities linked to cities’ economic vitality and creativity scale superlinearly with population size. This does not translate however in larger employment rates, as the number of employees scales linearly with population size.

### 8.2.4 *Human interactions*

At the heart of Bettencourt’s model [38] to explain the superlinear scaling of quantities associated with wealth and creativity is the behaviour of the total number of interactions between individuals with the size of the city. In an attempt to test this hypothesis, Schl  pfer et al. [203]

looked at the scaling of the cumulative number of contacts  $K$  that people had over the phone, using mobile phone data in Portugal, and landlines in the UK. They also looked at the cumulative call volume (total number of minutes called) and the cumulative number of calls, and found that the three quantities scale superlinearly with population size (see Table 4).

They further found that the number of non-returned calls showed a larger exponents than the number of calls, meaning that the number of solicitations an individual gets is greater in large cities.

Quantity	Exponent	City Definition
Cumulative phone contacts	1.12	Morphological (Portugal)
Cumulative phone contacts	1.13	Administrative (Portugal)
Cumulative call volume	1.11	Morphological (Portugal)
Cumulative call volume	1.15	Administrative (Portugal)
Cumulative number of calls	1.10	Morphological (Portugal)
Cumulative number of calls	1.13	Administrative (Portugal)

Table 4: **Interactions over the phone.** Scaling of the cumulative number of phone contacts, phone calls and the cumulative call volume over 409 days in Portugal. As for the scaling of the surface area, administrative and morphologically defined cities exhibit similar exponents. The scaling for LUZ (european functional definition) shows a behaviour compatible with a linear scaling, although the number of points (9) is not large enough to conclude. The data were obtained from a mobile phone provider, and all quantities are rescaled to take into account the variation of the operator’s coverage between cities.

#### 8.2.5 *Mobility of individuals, and environmental impact*

Because cars are widely used (at least in the US), and because peak travel demand on the roads corresponds to journey-to-work trips, most of the information available on the mobility of individuals concerns the commuting to work, often by car.

Samaniego and Moses [199] showed that the total number of miles driven in US Urban Areas (morphological definition) rescaled by the total surface area scales sublinearly with population size, with a non-trivial exponent (that is, different from  $1/2$ . More details in the next chapter). We showed in a later study [4] that the total distance driven scales linearly with population size in Urban Areas. Also related to commuting, and the use of personal vehicles, is the evolution of the total consumption of gasoline with city size. Bettencourt et al. showed that gasoline sales in Metropolitan Statistical Areas scaled sublinearly with population size [39] (see Table 5 for values).

Hopefully, new data such as mobile phone data should be able to inform us about other trips, which represent no less than 80% of all trips undertaken in the United States! [201].

A diseconomy associated with the mobility of individuals is the quantity of CO<sub>2</sub> emitted due to transportation (and polluting substances). Using different city definitions, different authors find very different behaviours. The authors of [88] find that transport-related CO<sub>2</sub> emissions in Metropolitan Statistical Areas in the US scale sublinearly with population size, while the authors of [3, 171] find that they scale superlinearly with population size for US Urban Areas (morphological definition). We will come back to this in the next Chapter.

Quantity	Exponent	City Definition	Study
Distance driven	1	Morphological (US)	Louf & Barthelemy [4]
Gasoline sales	0.79	Functional (US)	Bettencourt et al. [39]
CO <sub>2</sub> emissions	1.42	Morphological (US)	Oliveira et al. [171]
CO <sub>2</sub> emissions	1.37	Morphological (US)	Louf & Barthelemy [5]
CO <sub>2</sub> emissions	0.93	Functional (US)	Fragkias et al. [88]

Table 5: **Mobility.** Scaling relationships linked to the individual mobility in cities. The three scaling exponents regarding the CO<sub>2</sub> emissions *due to transportation* were obtained using the Vulcan data (<http://vulcan.project.asu.edu/>) which provide measurements of the CO<sub>2</sub> emissions on a 10 km  $\times$  10 km grid. The difference between the three studies is in the method used to delineate cities: Fragkias et al. [88] rely on the Metropolitan Statistical Areas defined by the Census Bureau, Oliveira et al. [171] rely on the City Clustering Algorithm [196] (morphological criterion) while we rely on the Urban Areas defined by the Census Bureau.

#### 8.2.6 Basic commodities

We can also wonder how the consumption of basic commodities (housing, water, electricity) per capita changes with population size. By far the most expected result, Bettencourt et al. showed [39] that the total water consumption (in China), the total electrical consumption (in China), and the total housing (in the US) are proportional to the population (see Table 6).

#### 8.2.7 Infrastructure

What about infrastructure, and the alledged economies of scale? Do we need to build less roads, lay less cables for every individual in larger cities? This question can be answered by looking at the scaling of the

Quantity	Exponent	City definition	Study
Total housing	1.00	Functional (US)	Bettencourt et al. [39]
Total electrical consumption	1.05	Administrative (China)	Bettencourt et al. [39]
Total water consumption	1.01	Administrative (China)	Bettencourt et al. [39]

Table 6: **Basic commodities.** Scaling of the total housing, electrical consumption and water consumption with population size. All exponents are compatible with a linear behaviour (within the 95% confidence interval error bars).

length of roads, cables, etc. in cities: if the exponent is smaller than one, larger cities need less infrastructure per capita.

Veregin and Tobler, using the 1980 US Census DIME files (a lot less convenient to use than shapefiles!) showed that the number of street segments—the portion of road between two intersections—scaled sublinearly with the size of urban areas [223] (see Table 7).

Arguably, the total length of the street network is more relevant to measure costs in terms of infrastructure. In [4], we provide evidence for the sublinear scaling of total street length with the population size of urban areas (Table 7).

Finally, Bettencourt et al. showed that the length of electric cables in German cities scaled sublinearly with population size [39]. So far, studies thus indicate that cities indeed realise some economies of scale.

Quantity	Exponent	City definition	Study
Street segments	0.83	Morphological (US)	Veregin & Tobler [223]
Street length	0.86	Morphological (US)	Louf & Barthelemy [4]
Electric cables length	0.87	Administrative (Germany)	Bettencourt et al. [39]

Table 7: **Infrastructure.** Scaling of the total number of street segments, the total length of roads and the total length of electrical cables of cities as a function of population. The three quantities exhibit a sublinear scaling behaviour, implying that larger cities need less infrastructure per capita, thereby realising economies of scale.

### 8.3 SUMMARY

The above review of the literature begs several questions.

First, most of the scaling exponents that are found in the literature (all but linear scalings) are highly non-trivial, in the sense that their values seem somewhat arbitrary. We argued at the beginning of this Chapter that these exponents were the signature of the processes happening within cities. But it is not clear what mechanisms can lead to these values. In the following Chapter, we will provide a model that

reproduces the exponents observed on quantities that are related to the mobility of individuals.

A second issue has to do with the fact that studies find different exponent for the exact same quantities. The problem does not lie so much with the numerical differences, but in the qualitative difference: some quantities are found to scale sublinearly in a context, and superlinearly in another. For instance, the  $\text{CO}_2$  emissions scale differently with population size in different studies. While studies focusing on Urban Areas or equivalent (in the US) find that emissions scale superlinearly with population size [4, 171], studies interested in Metropolitan Statistical Areas report a sublinear scaling [88]. This calls for an explanation that we will sketch in Chapters 9 and 10.



*I remember my friend Johnny von Neumann used to say,  
‘with four parameters I can fit an elephant  
and with five I can make him wiggle his trunk’*

— Enrico Fermi (quoted in [76])

A common trait shared by all complex systems – including cities – is the existence of a large variety of processes occurring over a wide range of time and spatial scales. The main obstacle to the understanding of these systems therefore resides in uncovering the hierarchy of processes and in singling out the few ones which govern their dynamics. Albeit difficult, the hierarchisation of processes is of prime importance. A failure to do so leads to models which are either too complex to give any real insight into the phenomenon, or too simple and abstract to have any resemblance with reality. As a matter of fact, despite numerous attempts [91, 144, 26, 89, 37, 38], a theoretical understanding of many observed empirical regularities in cities is still missing.

Here we show that the spatial structure of the mobility pattern controls the scaling behaviour of many quantities in urban systems. Indeed, cities are not only defined by the spatial organisation of places fulfilling different functions – shops, places of residence, workplaces, etc. – but also by the way individuals move among them. Understanding where people live, where and how they travel within the city thus appears as a necessary step towards a scientific theory of cities.

## 9.1 A NAIVE APPROACH

We start by presenting some naive arguments to estimate the scaling exponents for the area  $A$ , the total daily distance driven  $L_{tot}$  and the total lane miles  $L_N$ . Although these predictions turn out to be wrong, naive scalings are useful as a first approach to the problem as they allow us understand how the different quantities relate to one another.

### 9.1.1 Surface area

We first would like to estimate the dependence of the area  $A$  of a city on its population  $P$  – a long standing problem in the field [214, 29].

NAIVE ARGUMENT. A first crude approach is to assume that cities evolve in such a way that their population density  $\rho = P/A$  remains constant. This assumption immediately implies that the area should scale linearly with population

$$A \sim \lambda^2 P \quad (44)$$

where  $\lambda^2$  is the average surface occupied by each individual (the assumption of a constant density is then equivalent to the one of a constant average surface per capita).

REALITY. The naive argument does not compare well with reality. We plot the scaling of the surface area versus population for US Urban Areas on Figure 17. A fit assuming a power-law dependence gives an exponent

*All  $\pm$  intervals are  
95% confidence  
intervals.*

$$\beta_A = 0.85 \pm 0.01 \quad (r^2 = 0.93) \quad (45)$$

A result which agrees with previous measurements made on morphologically defined cities (see [29] or Chapter 8). This means that the average surface occupied by each individual decreases with city size. Or equivalently, that the population density increases with city size. The prediction given by the naive model is therefore quantitatively – and worse, qualitatively – different from the behaviour observed empirically.

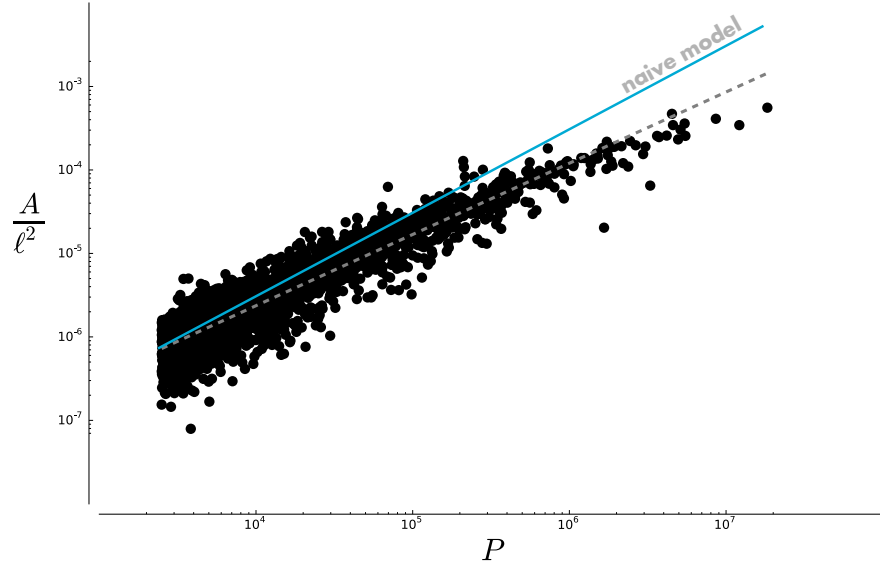


Figure 17: **Spatial footprint.** Scaling of the surface area of US urban areas with population size, and what would be expected with a naive model (blue solid line). A fit assuming a powerlaw dependence (dashed line) gives an exponent  $\beta_A = 0.85 \pm 0.01$  ( $r^2 = 0.93$ ).



### 9.1.2 Total length of road

NAIVE MODEL. We would now like to estimate the total length  $L_N$  of all the roads within a city. If we consider that the network formed by streets is such that all the nodes (intersections) are connected to their closest neighbour, the typical length of a road segment is given by

$$\ell_R \sim \sqrt{\frac{A}{N}} \quad (46)$$

where  $N$  is the number of intersections [23]. Previous studies of road networks in different regions, and over extended time periods [218, 24], have shown that the number of intersections is proportional to the population size. Therefore, the typical length of a road segment (between two intersections) varies with the population size  $P$  as

$$\ell_R \sim \sqrt{\frac{A}{P}} \quad (47)$$

and the total length of the network  $L_N \sim P\ell_R$  should then scale as

$$\frac{L_N}{\sqrt{A}} \sim \sqrt{P} \quad (48)$$

Using the naive scaling for the dependence of  $A$  on population size given previously in Eq. 44 we finally get

$$L_N \sim P \quad (49)$$

REALITY. Again, the naive argument does not compare well with reality. We fit the data for US Urban Areas (see Figure 18) assuming a powerlaw dependence and find an exponent

$$\boxed{\beta_R \sim 0.765 \pm 0.033 \ (r^2 = 0.92)} \quad (50)$$

Note that the relation between the length and the number of nodes given by Eq. 47, as well as the relation between number of intersections and population, have been verified independently in the literature. The observed discrepancy on the exponent of  $L_N$  is therefore certainly due to the scaling of the surface area.

### 9.1.3 Total commuting distance

The total commuting distance  $L_{tot}$  is determined by two different constraints. First the individual constraint: individuals make the decision

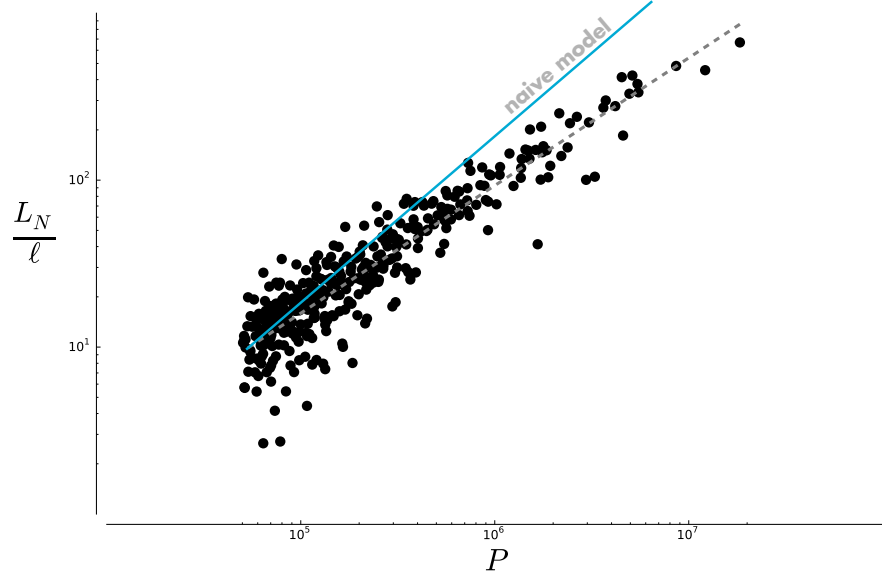


Figure 18: **Length of roads.** Scaling of the total length of roads in US Urban Areas versus the total population. A fit assuming a powerlaw dependence (dashed grey line) gives an exponent  $\beta_R \sim 0.765 \pm 0.033$  ( $r^2 = 0.92$ ). The behaviour is qualitatively different from what would be expected with a naive model (solid blue line).

about where they are going to live and work; they have their own behaviour and limitations. However, the individuals' choices are also limited by the city structure itself, that is by the respective distributions of jobs and residences across the city.

#### 9.1.3.1 Influence of the individual constraint

The first constraint on the commuting distance comes from individuals' limitations and behaviour. We make here the simple assumption that individuals choose their residence and work place such that their total commuting distance is fixed (or at least, is smaller than a certain value) and equal on average to  $\ell_C$ . In that case, we would simply have

$$\frac{L_{tot}}{P} \sim \text{constant} = \ell_c \quad (51)$$

(by constant, we mean independent from the population size of the city). As surprising as it may seem, the data show that  $L_{tot}/P$  can indeed be considered independent from  $P$  (with a value of approximately 23 miles for the US, see Figure 19), in agreement with the individual constraint assumption (Eq. 51). This finding is also in agreement with the results drawn from census data in Germany by [227]. This does not mean, of course, that the distance driven is the same for every city. As one can see on Figure 19, the fluctuations are quite important between cities.

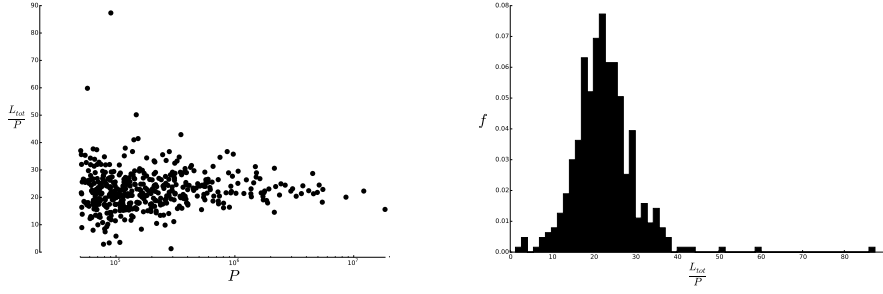


Figure 19: **Commuting distance & individual choice.** Constant daily driven distance per capita. (a) daily total driven distance per capita as a function of population for 441 urbanised area in the US in 2010. The data shown in the plot are compatible with a population-independent behaviour. (b) Histogram of the daily total driven distance per capita for the same cities. The average daily driven distance is 23 miles, and the standard deviation 7 miles.

### 9.1.3.2 Influence of the city structure

The easiest way to understand the influence of the city constraints is to consider two limiting cases: the totally centralised (monocentric) city where everyone goes to work to a single center, and the totally decentralised city where everyone goes to work to the nearest location (see Figure 20) [199].

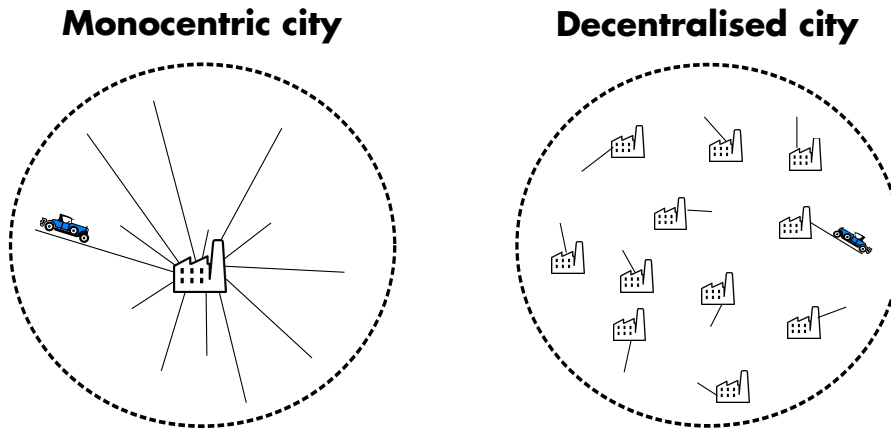


Figure 20: **Limiting cases.** Representation of the monocentric city (left) and the totally decentralised city (right), two extreme models for the shape of mobility patterns.

**MONOCENTRIC.** If we first assume that the city is monocentric, individuals are all commuting to the same center and the typical com-

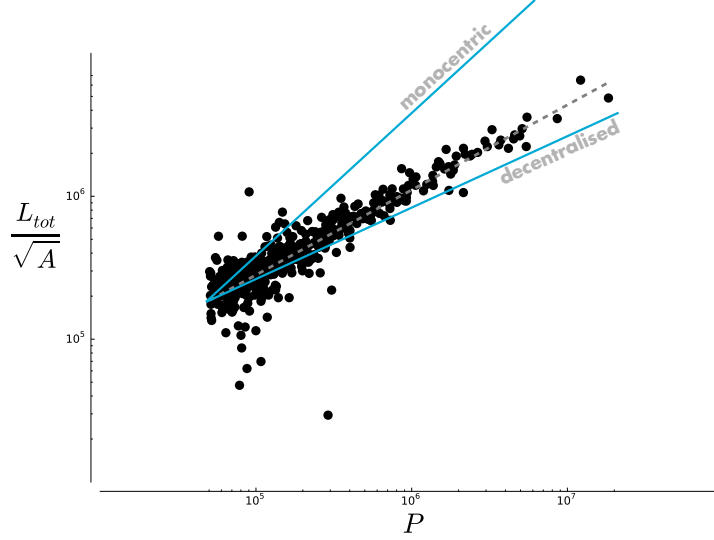


Figure 21: **Commuting distance & city structure.** Scaling of the total yearly commuted distance normalised by the city’s surface area with population size for US Urban Areas. The blue lines show the behaviours that would be expected for a monocentric and a totally decentralised city. The dashed line represents the fit assuming a powerlaw dependence, which yields an exponent  $\beta = 0.595 \pm 0.026$  ( $r^2 = 0.90$ ).

muting distance  $\ell_c^m$  is controlled by the typical size of the city of order  $\sqrt{A}$ , so that

$$\frac{L_{tot}^m}{\sqrt{A}} \sim P \quad (52)$$

DECENTRALISED. On the other hand, if we assume that the city is completely decentralised, the typical commuting distance is of order the nearest neighbour distance  $\sqrt{A}/\sqrt{P}$ , and we obtain

$$\frac{L_{tot}^d}{\sqrt{A}} \sim \sqrt{P} \quad (53)$$

REALITY. The scaling of the total driven distance for Urban Areas (morphological definition) is shown on Figure 21, and the exponent sits between the ones of the monocentric and decentralised cities

$$\boxed{\beta_L = 0.595 \pm 0.026 \ (r^2 = 0.90)}$$

This comes as another evidence – different from that presented in Chapter 5 – that cities do not have a strictly monocentric structure. This result casts some further doubts about the model by Bettencourt [38] which implicitly assumes that cities are monocentric.

So far, so good. But how can we understand the non-trivial exponent that is observed? This is where the limiting case are helpful: if the exponent sits between the ones that would be obtained in a monocentric or decentralised city, surely, cities must adopt an intermediate structure.

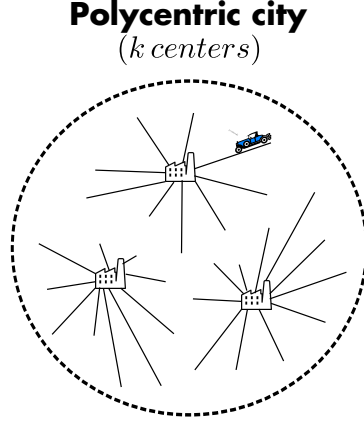


Figure 22: **Polycentric structure.** City with a polycentric structure, intermediate between the monocentric and totally decentralised situations.

One candidate stands out: the polycentric city (see Figure 22). Let us thus consider a polycentric city with  $k$  employment centers. The typical distance commuted by individuals is then given by

$$\ell_c \sim \sqrt{\frac{A}{k}} \quad (54)$$

So that

$$\frac{L_{tot}}{\sqrt{A}} = \frac{P}{\sqrt{k}} \quad (55)$$

Therefore if, as we showed in the previous part, the number of centers increases sublinearly with population, we would have a scaling of the form  $L_{tot}/\sqrt{A} \sim P^{\beta_L}$  where  $\beta_L \in [1/2, 1]$ . The previous expression is consistent with that of  $A/\lambda^2$  and  $L_{tot}/P$  if

$$\beta_L = 1 - \frac{\beta_A}{2} \quad (56)$$

which is indeed what we observe empirically (up to error bars). We conclude from this preliminary empirical analysis that, in order to compute the various exponents, we need to better describe the structure of commuting patterns. In other words, we need to find a description of cities that goes beyond the naive monocentric or totally decentralized views, and which accounts for the observed sub-linear scaling of the surface area  $A$ .

Quantity	Naive exponent	Measured value
$A$	1	0.85 ( $r^2 = 0.93$ )
$L_N/\sqrt{A}$	0.5	0.42 ( $r^2 = 0.83$ )
$L_N$	1	0.89 ( $r^2 = 0.77$ )
$L_{tot}/\sqrt{A}$	$\{0.5, 1\}$	0.60 ( $r^2 = 0.90$ )
$L_{tot}/P$	1	0.03 ( $r^2 = 0.04$ )

Table 8: **Naive exponents and measured values.** This table displays the value of the exponent governing the behavior with the population  $P$  obtained by naive arguments and the value obtained from empirical data. The discrepancies reveal the failure of the naive scaling arguments and the necessity to go further and model mobility patterns.

## 9.2 BEYOND NAIVE SCALINGS: MODELING THE MOBILITY PATTERNS

The previous results, in particular the behaviour of the total commuting length with population, hint at the necessity to better describe the structure of the mobility patterns (Table 8). This is exactly what the model presented in the previous chapter does.

Using the relation that we derived for the number of centers, we will see how we can understand the values of the exponents presented earlier in this chapter. We will also see how the model allows us to understand the scaling of other quantities, namely the total time spent in traffic and the total  $\text{CO}_2$  emissions due to transportation.

### 9.2.1 Area

According to the model introduced in Chapter 6, the number of centers is a function of population and the area

$$k = F(A, P) \tag{57}$$

and we need an additional equation in order to get a closed system. Here we focus on the area and its evolution with the population size, which reflects the growth process of the city.

In the following, we will investigate two different approaches. It is worth noting that both approaches give results in qualitative agreement, showing that some stylized facts —such as super- or sublinearity— are very robust.

FITTING PROCEDURE. In the absence of knowledge of the processes responsible for urban sprawl, we can assume that the area behaves as

$$A \sim P^a \quad (58)$$

where  $a$  is the exponent to be determined by fitting data. The empirical value for the exponent for the US data is  $a \simeq 0.85$ . Once this exponent is given we can then compute the various exponent for the quantities of interest. We get for the number of centers  $k$

$$k \sim P^{\frac{\mu+a/2}{\mu+1}} \quad (59)$$

which is sublinear as long as  $a < 2$ , in agreement with the empirical results for US cities. As we will see, this approach yields the same qualitative behaviours as those predicted with the method of the next section. In other words, even if the main mechanism behind urban sprawl is not congestion, the conclusions of this paper are not affected as long as the area scales *sublinearly* with population.

COHERENT GROWTH. Let us now assume that the scaling of  $A$  with population is determined by the number of activity centers and the constant commuting length of individuals. This means that the growth of the area is controlled by the appearance of new activity centers.

If we assume that a city is organized around  $k$  activity centers and that the attraction basin of each of these centers are spatially separated [2] (See on Figure 22), we then have  $A \sim k A_1$  where  $A_1$  is the area of each subcenter's attraction basin. This area  $A_1$  is related to the average individual commuting distance by  $\sqrt{A_1} \sim L_{tot}/P$ , and we obtain

$$A \sim k \left( \frac{L_{tot}}{P} \right)^2 = k \ell_c^2 \quad (60)$$

This leads to expression for the number of centers

$$k \sim P^{\frac{2\mu}{2\mu+1}} \quad (61)$$

which is always smaller than 1, also in agreement with the empirical results for US cities. We can now also compute the scaling of the surface area

$$\frac{A}{\ell_c^2} \sim \left( \frac{P}{c} \right)^{\frac{2\mu}{2\mu+1}} \quad (62)$$

We further assume that  $L_{tot}/P$  is a fraction of the longest possible journey  $\ell$  individuals can afford, that is to say

$$\ell_c \sim \ell \quad (63)$$

It is important to note that if  $\ell_c$  is independent from  $\ell$ , the quantitative predictions of our model would still hold.

The final expression for the area is then here given by

$$\frac{A}{\ell^2} \sim \left(\frac{P}{c}\right)^{2\delta} \quad (64)$$

where  $\delta = \frac{\mu}{2\mu+1}$ . The exponent  $\delta$  is smaller than  $1/2$  whatever  $\mu \geq 0$ , which implies that the surface area of cities increases *sublinearly* with population. In other words, the density of cities *increases* with population. This prediction is verified with data about land area of urbanized areas in the US (Figure 17). We find  $\beta_A = 0.85 \pm 0.01$  which is not too far from the theoretical value  $2\delta_{th} = 0.64 \pm 0.12$ , equal to  $\alpha$  in this case.

Because the area of a city results from centuries of evolution, we do not a priori expect our model – where individual vehicles are assumed to be the only vector of mobility – to give a prediction valid for all countries and all times. Nevertheless, these results give us reasons to believe that the spatial structure of the journey-to-work commuting might be the dominant factor in the dependence of land area on population. In the following, we will use the above numerical value to compute other scaling exponents.

### 9.2.2 Total commuting distance

Using Eq. 51 and Eq. 64 we are now able to compute  $L_{tot}/\sqrt{A}$

$$\frac{L_{tot}}{\sqrt{A}} = P \left(\frac{P}{c}\right)^{-\delta} \quad (65)$$

We plot  $L_{tot}/\sqrt{A}$  for urbanized areas in the US on Figure 21, and one can verify in Table 9 that the exponent predicted from the previously measured value of  $\alpha$  agrees well with the exponent measured on the data.

### 9.2.3 Total length of roads

If we use the previously derived expression for the area  $A$ , we find

$$L_N \sim \ell \sqrt{P} \left(\frac{P}{c}\right)^{\delta} \quad (66)$$



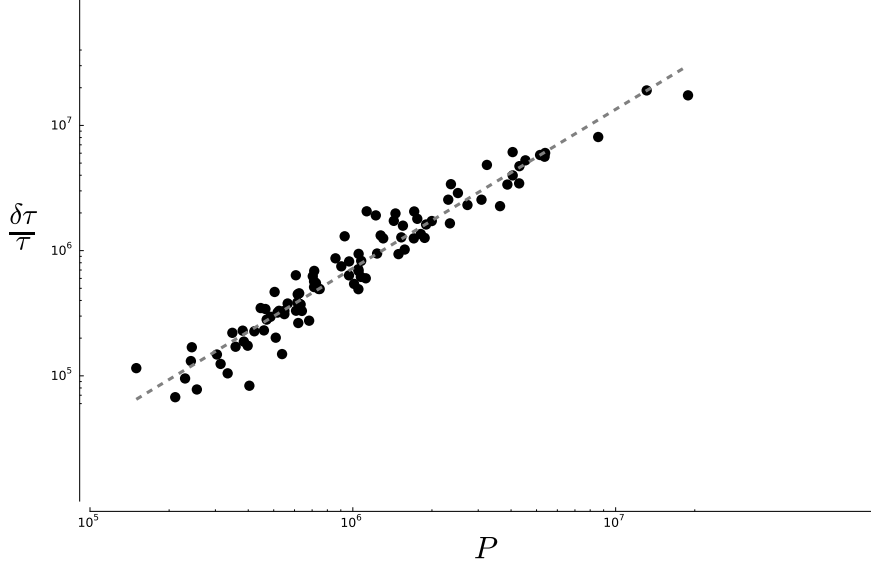


Figure 23: **Congestion and delay.** Scaling of the total delay due to congestion of US urban areas with population size. A fit assuming a powerlaw dependence of the total delay on population size yields an exponent  $\beta_D = 1.270 \pm 0.067$  ( $r^2 = 0.97$ ).

The quantity  $\delta$  is less than  $1/2$ , which implies that  $L_N$  scales *sublinearly* with the city's population size. In other words, larger cities need less roads per capita than smaller ones: we recover the fact that the agglomeration of people in urban centers involves economies of scale for infrastructures.

#### 9.2.4 Total delay due to congestion

Unfortunately, the agglomeration of activities in cities does not only generate economies. Congestion, for instance, is a major diseconomy associated with the concentration of people in a given area. A simple way to quantify the impairment caused by traffic congestion is through the total delay it generates. If we make the first order approximation that the average free-flow speed  $v$  is the same for everyone, the total delay due to congestion is given –according to our model– by

$$\delta\tau = \frac{1}{v} \sum_{i,j} d_{ij} \left( \frac{T_j}{c} \right)^\mu \quad (67)$$

If we assume that all the centers share the same number of commuters – a reasonable assumption within the model presented in Chapter 6 [2] – we obtain

$$\delta\tau \sim \frac{L_{tot}}{v} \left( \frac{P}{k} \right)^\mu \quad (68)$$

which, using the expressions for  $L_{tot}$  and  $A$  given in Eq. 65 and Eq. 64 respectively, gives

$$\delta\tau \sim \frac{\ell P}{v} \left( \frac{P}{c} \right)^\delta \quad (69)$$

The total commuting time corresponding to the same distance but without congestion scales as  $\tau_0 \sim L_{tot}$  and thus less rapidly than the total delay which scales *super-linearly* with population (even when polycentricity is taken into account). This means that, for the largest cities, delays due to congestion actually dominate the time spent in traffic, and that economical losses *per capita* due to the time lost in congestion –and the corresponding strain on people’s life– increase with the size of the city.

The prediction  $1 + \delta = 1.32$  agrees well with the empirical measure (see Table 9 and Figure 23)

$$\boxed{\beta_D = 1.270 \pm 0.067 \ (r^2 = 0.97)} \quad (70)$$

#### 9.2.5 Transport related CO<sub>2</sub> emissions

Another diseconomy associated with congestion is the quantity of CO<sub>2</sub> emitted by cars and the gasoline consumed by motor vehicles. This amount not only depends on the distance that has been driven, but also on the traffic during the journey. It indeed turns out that for the same length driven, a car burns more oil when the traffic is heavy than when the road is clear. Within our model, the presence of traffic is seen in the time spent to cover a given distance, and we write that the quantity of CO<sub>2</sub> emitted by a vehicle is proportional to the total time spent in traffic, leading to

$$Q_{CO_2} = q \sum_{ij} d_{ij} \left[ 1 + \left( \frac{T_j}{c} \right)^\mu \right] \quad (71)$$

where  $q$  is the average quantity of CO<sub>2</sub> produced per unit time. In the polycentric case with  $k = k(P)$  subcenters, the typical trip length  $\overline{d_{ij}}$  is given by  $\sqrt{A/k}$  and we obtain

$$Q_{CO_2} = q \ell P \left[ 1 + \left( \frac{P}{c} \right)^\delta \right] \quad (72)$$

The first term in brackets is a constant, and the quantity of CO<sub>2</sub> is thus dominated by congestion effects at large populations

$$Q_{CO_2} \sim q \ell P \left( \frac{P}{c} \right)^\delta \quad (73)$$

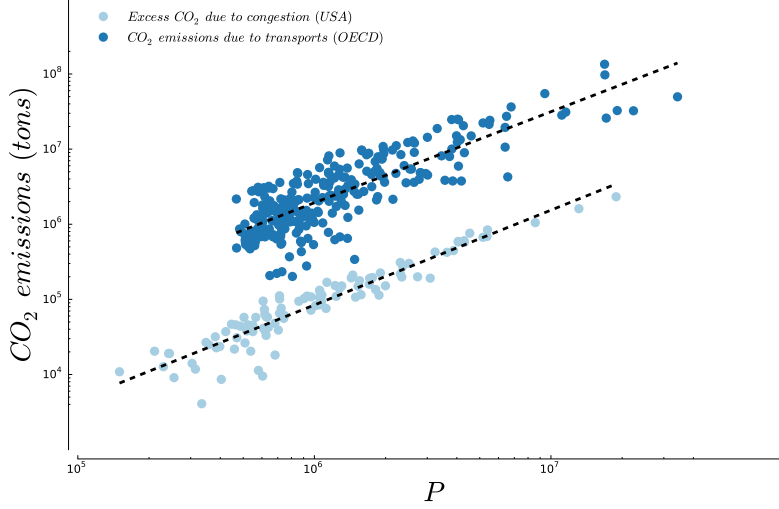


Figure 24: **Congestion and CO<sub>2</sub> emissions.** Variation of CO<sub>2</sub> emissions due to transport with city size. In blue, excess CO<sub>2</sub> (in tons) due to congestion, as given by the Urban Mobility Report (2010) for 101 metropolitan areas in the US. In green, we show the estimated CO<sub>2</sub> emissions (in tons) due to transports, as given by the OECD for 268 metropolitan areas in 28 different countries. The dashed yellow lines represent the least-square fit assuming a power-law dependency with multiplicative noise, which gives respectively  $Q_{\text{CO}_2} \sim P^{1.262 \pm 0.089}$  ( $r^2 = 0.94$ ) for the US data and  $Q_{\text{CO}_2} \sim P^{1.212 \pm 0.098}$  ( $r^2 = 0.83$ ) for the OECD data.

and the total daily transport-related CO<sub>2</sub> emission per capita thus scales as

$$\frac{Q_{\text{CO}_2}}{P} \propto q\ell \left(\frac{P}{c}\right)^\delta \quad (74)$$

The quantity of CO<sub>2</sub> emitted per capita in cities thus increases with the size of the city, a consequence of congestion. This prediction agrees with the exponent we measure (Figure 24) on data gathered for US and OECD cities (see Table 9)

$$\boxed{\beta_C = 1.262 \pm 0.089 \ (r^2 = 0.94)} \quad (75)$$

### 9.3 DISCUSSION

#### 9.3.1 Travel-time budget and congestion

The total commuting time  $T$  can be written as

$$T = \tau_0 + \delta \tau \quad (76)$$

where  $\tau_0 = L_{tot}/v \sim P$  is the free-flow commuting time and  $\delta\tau \sim P^{1+\delta}$  the excess commuting time computed above. The first thing we notice when looking at the respective population dependence of both quantities, is that, in large cities, the total commuting time is dominated by the time spent in congestion. Indeed, we have

$$\frac{T}{\delta\tau} \xrightarrow{P \gg 1} 1 \quad (77)$$

Which agrees with one's (at least our) experience of driving in large cities.

The second remark is linked to a long-standing belief in the study of urban systems that individuals possess a constant travel-time budget [232]. We can easily see, however, that this hypothesis is wrong. Indeed, in the limit of large cities, the individual commuting time is given by

$$\frac{\delta\tau}{P} \sim P^\delta \quad (78)$$

In other words, the *individual commuting time increases with the size of the city*. Note that not only is this a consequence of the model, but also of the data analysis (see Figure 23). The constant travel-time budget hypothesis is thus refuted. The reason for the discrepancy between previous measures and our results comes from the fact that these studies considered averages over large regions, rather than averages at the city level.

### 9.3.2 Newman & Kenworthy

The consumption of gasoline is proportional to the emission of  $\text{CO}_2$  and the time spent driving. The total daily gasoline consumption is thus given by

$$Q_{gas} \sim q \ell P \left( \frac{P}{c} \right)^\delta \quad (79)$$

where  $q$  is the average quantity of gasoline needed per unit time. From this expression, we see that the total daily gasoline consumption per capita scales as

$$\frac{Q_{gas}}{P} \sim \ell \sqrt{\frac{P}{\rho}} = \ell \sqrt{A} \quad (80)$$

and is therefore not a simple function of population density, in contrast with what was suggested by the seminal paper of Newman and

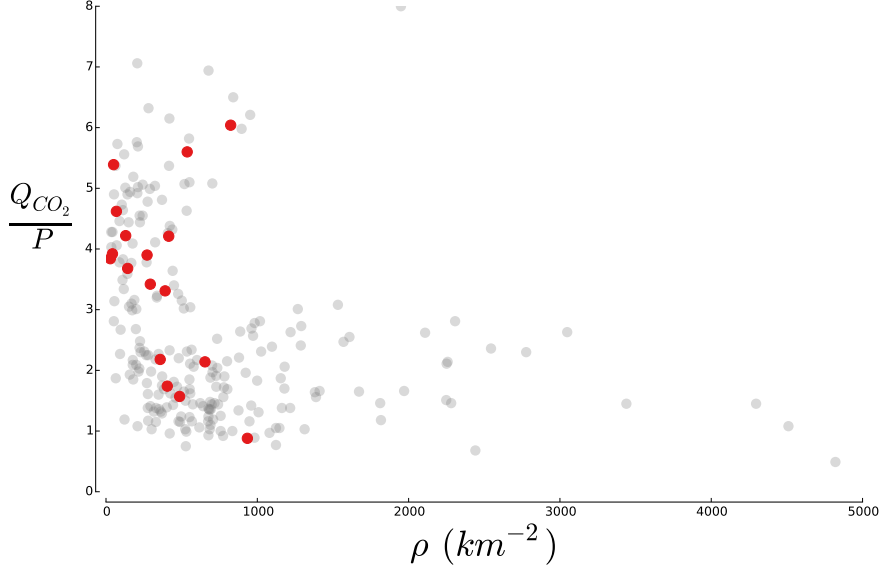


Figure 25: **Newman & Kenworthy.** Per capita  $\text{CO}_2$  emissions versus the population density of cities belonging to OECD countries. The cities also present in the Newman & Kenworthy dataset are represented in red. This curve casts serious doubt on the fact that energy consumption is a simple function of density.

Kenworthy [165]. We plot on Figure 25, the average individual  $\text{CO}_2$  emissions (used as a proxy for gasoline consumption) as a function of the density for OECD cities. The points corresponding to cities that were in the original study [165] are highlighted. The relation is a lot less clear than the one presented originally.

We then plot the same quantity as a function of  $\sqrt{A}$ , the prediction given by Eq. 80, on Figure 26. As one can see, the prediction is far from perfectly followed. If anything, this figure, combined to Figure 25 show that the debate, in the absence of a clear-cut conclusion, is not over. At this stage, more data about gasoline consumption – preferably for cities belonging to the same system of cities – is needed to explore this prediction.

### 9.3.3 *Monocentric versus polycentric*

Although polycentricity emerges naturally from our model as a result of congestion, many circumstances can prevent or foster the appearance of new activity centers in a city. There are many debates as to whether policies should favour polycentric or monocentric development of cities. Most of them are based on ideologies and opinions about how cities should be, very few are based on a quantitative understanding of the city as a complex system. Although this only represents a small part of the debate, our model allows to quantify the effect of polycentricity on the total delay due to congestion.

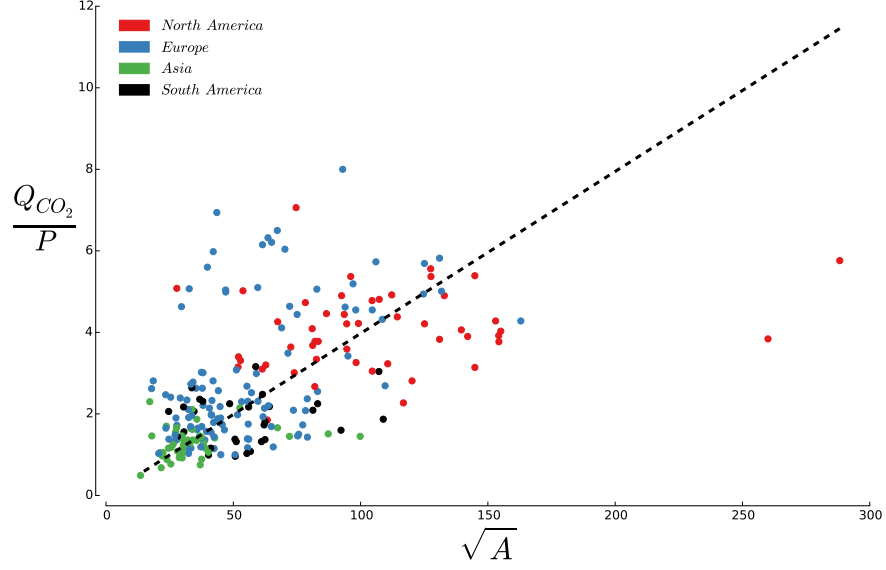


Figure 26: **Newman & Kenworthy revisited.** Per capita  $\text{CO}_2$  emissions versus  $\sqrt{A}$  for cities of countries that belong to the OECD. The dashed line represents the obtained linear fit, as predicted by Eq. 80 ( $r^2 = 0.55$ ). The agreement is poor, which may be due to the fact that cities all belong to different systems of cities (and thus have a different prefactor).

We can indeed compute the total delay due to congestion in the case of a monocentric configuration. In this situation, all the population commutes to a single destination 1 and we have

$$\delta\tau_{mono} = \frac{1}{v} \sum_i d_{i1} \left(\frac{P}{c}\right)^\mu = L_{tot} \left(\frac{P}{c}\right)^\mu \quad (81)$$

It follows, using the expression given above for  $L_{tot}$

$$\delta\tau_{mono} = \frac{\ell}{v} P^{1+\mu} \quad (82)$$

From the fact that  $1 + \mu > 1 + \frac{\mu}{2\mu+1}$ , we indeed find that the total delay due to congestion is worse for monocentric cities than it is for polycentric cities with the same population, which agrees with the usual intuition. More precisely the ratio of delays is given by

$$\frac{\delta\tau_{mono}}{\delta\tau_{poly}} \sim \left(\frac{P}{c}\right)^\beta \quad (83)$$

where the exponent is of order  $\beta \approx 0.57$ . Therefore, even though diseconomies associated with polycentric cities scale superlinearly with population, it would be even worse if we did not let cities evolve from the

monocentric situation. The same reasoning applies to the consumption of gasoline and the  $\text{CO}_2$  emissions.

This suggests that, everything else being equal, polycentricity should be favoured for quality of life and environmental reasons.

Quantity	Theoretical expression ( $\delta = \alpha/\alpha + 1$ )	Predicted exponent	Measured value
$L_{tot}$	$P$	1	$1.03 \pm 0.03$ ( $r^2 = 0.95$ )
$A/\ell^2$	$(\frac{P}{c})^{2\delta}$	$2\delta = 0.78 \pm 0.20$	$0.853 \pm 0.011$ ( $r^2 = 0.93$ )
$L_N/\ell$	$\sqrt{P} (\frac{P}{c})^\delta$	$\frac{1}{2} + \delta = 0.89 \pm 0.10$	$0.765 \pm 0.033$ ( $r^2 = 0.92$ )
$\delta\tau/\tau$	$P (\frac{P}{c})^\delta$	$1 + \delta = 1.39 \pm 0.10$	$1.270 \pm 0.067$ ( $r^2 = 0.97$ )
$Q_{gas,CO_2}/\ell$	$P (\frac{P}{c})^\delta$	$1 + \delta = 1.39 \pm 0.10$	$1.262 \pm 0.089$ ( $r^2 = 0.94$ )
			$1.212 \pm 0.098$ ( $r^2 = 0.83$ )
$L_N/\sqrt{A}$	$\sqrt{P}$	0.5	$0.42 \pm 0.02$ ( $r^2 = 0.83$ )
$L_{tot}/\sqrt{A}$	$P (\frac{P}{c})^{-\delta}$	$1 - \delta = 0.61 \pm 0.10$	$0.595 \pm 0.026$ ( $r^2 = 0.90$ )

Table 9: **Summary of the scaling exponents.** This table displays the predicted theoretical behavior and the empirical observations versus the population size  $P$  for different quantities:  $L_{tot}$  is the daily total driven distance,  $A$  is the area of the city,  $L_N$  is the total length of the road network,  $\delta\tau$  is the daily total delay due to congestion,  $Q_{gas}$  is the yearly total consumption of gasoline and  $Q_{CO_2}$  is the total  $\text{CO}_2$  emissions emitted yearly due to transportation. In the third column, we show the predicted values of the exponent of  $P$  using the value of  $\alpha$  measured on US employment data, and in the fourth column, the value of the exponents directly measured on data about US and OECD cities. The measured values are in good agreement with the prediction. In particular, the exponents for  $L_N$  and  $\delta\tau$  are consistent with our prediction that their difference should be  $1/2$ .

#### 9.3.4 Outlook

The superlinear increase of congestion delay with population, and thereby of gasoline consumption and of  $\text{CO}_2$  emissions, has terrible consequences on the economy, the environment, health and well-being. The outlook is nothing short of grim in our ever-urbanising world. As the proportion of human beings living in cities dramatically increases – the UN expects the world population to be 67% urban in 2050 – wages are likely to increase [39] but not enough to compensate for the negative effects of congestion. As a result, if the individual car stays the dominant transportation mode, cities will put more strain on people’s life, while acting as catalysts for the production of  $\text{CO}_2$  greenhouse gas, which is responsible for an overall increase of the planet’s temperature [174].

It is currently believed that advantages associated with living in a large city outweigh the costs. Our results reveal however the existence

*Estimates are given in the United Nations’ 2011 World Urban Prospects.*

of very rapidly growing problems such as congestion and  $\text{CO}_2$  emissions, which inevitably begs the question of the sustainability of large cities. It might be time to cut down considerably the use of individual vehicles, or to consider the possibility of living in smaller or medium sized cities: the infrastructure costs ( $L_N$ ) may be larger, but the impact on the environment ( $\text{CO}_2$  emissions) and on the well-being of people (delays in congestion) would be beneficial.

The most striking fact about the above results is that despite the appearance of complexity that is conveyed by cities, most of their structure can be explained by the very simple and universal desire for the best achievable balance between income and commuting costs. Our model unifies mobility patterns, spatial structure of cities and allometric scalings in a framework that can be built upon.



## INTERPRETATIONS AND IMPLICATIONS OF SCALING LAWS

---

*There are no facts, only interpretations.*

— Friedrich Nietzsche

Although allometric scaling relationships are a powerful tool to explore the behaviour of cities, there are several continuing controversies in the literature. First, about their interpretation: do these relationships say something about cities and the processes they host, or cities as they relate to one another in a system of cities? Second, recent studies [16, 3, 62] have shown that the measured exponents are very sensitive to the way cities are defined. What does it imply for the study of these scalings and, more generally, cities?

### 10.1 WHAT SCALING LAWS TELL US ABOUT CITIES

Scaling laws are, in essence, cross-sectionnal studies of cities. As opposed to dynamical studies where one would follow the evolution of individual cities over time, scaling laws tells us about the behaviour of an *ensemble* of cities at a give point in time. Throughout Chapters 8 and 9, we have implicitly assumed that scaling laws are the signature of phenomena occuring at the intra-urban level. This assumption, we call *evolution interpretation*, is however not completely obvious.

Maybe the easiest way to understand the issues posed by this interpretation is through the comparison with Biology, where allometric scaling laws are also widely used. The interpretation of allometric scaling laws in Biology is straightforward, because the compared organisms are independent. Consider, for instance, the scaling of the metabolic rate of animals with their body mass [224, 20]. The mass of a given elephant at a point in time  $t$  is not correlated to the mass of any other living creature in the world. Therefore, the scaling relationship can only be understood as resulting from the existence of similar processes in the growth of these different animals. Cities are different. They are part of a bigger system – the system of cities – and interact constantly with one another. People change residence, companies relocate, goods are shipped and money is transferred. Therefore, as argued by Denise Pumain [185], scaling laws can also be construed as reflecting the redistribution processes within this system of cities. We call this the *differentiation interpretation*.

### 10.1.1.1 *The evolution interpretation*

The evolution interpretation (Figure 27) has been widely adopted in the scaling literature [39, 38, 4] without ever being clearly stated, let alone justified. It is based on two assumptions. The first assumption is that cities in the dataset are different realisations of the same system. Thus, as stated in Chapter 8, looking at the scaling of various quantities with population size is a way to probe the system's internal processes.

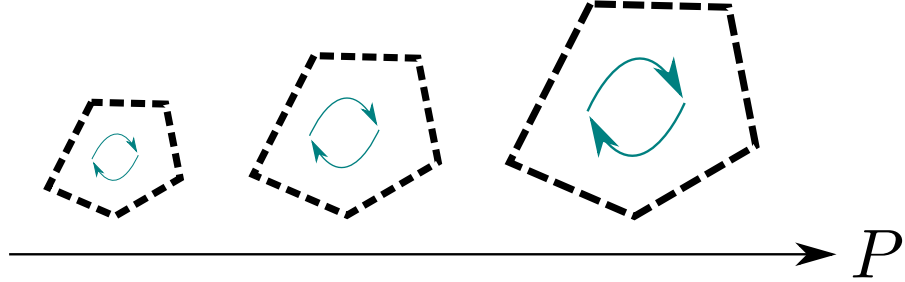


Figure 27: **Evolution interpretation.** In this interpretation we consider that cities are different realisations of the same system. The intra-urban processes – and the way they respond to population changes – are responsible for the non-linear scaling of the different quantities.

The second assumption has to do with the time scales over which the different processes occur. Indeed, if the processes responsible for the change in the value of the quantity  $Y$  being studied occur on timescales significantly larger than the timescale over which the population size changes, we cannot be sure the exponent value actually reflects the internal processes at the time we measure it. For instance, an abrupt increase in population size is not likely to be immediately reflected in the length of streets, while the evolution of the total commuted length will be almost instantaneous. In practice, the rate of population change in cities is small enough for the processes to follow, or the amplitude small enough for the induced error to be insignificant. Hence the observed stability in the value of some exponents.

The previous discussion has several important consequences. First, it hints at the difficulty to interpret the values of the observed deviations to scaling laws [40]. It is indeed difficult to assess to what extent deviations account for a real over- or under-performance of the city compared to the other cities, or for the time it takes for the studied quantity to react to population changes. Worse, the delayed adjustment to population changes introduces an irreducible uncertainty in the numerical values of the exponents themselves. Thus, the real error on the measured value of the exponent is very likely larger than what is usually indicated by the statistical error bars. Unfortunately, we cannot get a better estimate of the error until we understand in details the mechanisms responsible for the time evolution of the corresponding quantities. Until then, we should focus on (1) trying to understand the qualitative behaviour,

more than the exact numerical value of the exponents (2) be wary of interpreting exponent values that are close to 1 (typically between 0.90 and 1.10); in the absence of an alternative mechanistic explanation, the linear relationship has to be favoured due to its simplicity.

### 10.1.2 *The differentiation interpretation*

As Denise Pumain judiciously claims [187, 185], the evolution interpretation is not the only possible interpretation for scaling laws. In some cases and the mechanisms responsible for scaling relationships should be sought after in the hierarchical organisation of cities and their interactions.

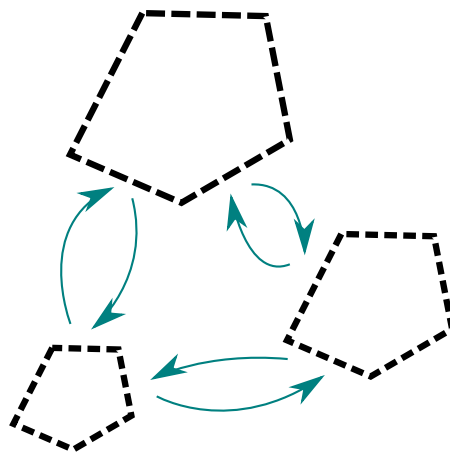


Figure 28: **Differentiation interpretation.** In this interpretation, we consider that the redistribution processes occurring within systems of cities are responsible for the non-linear scaling of quantities with city size in this system.

We briefly mentioned in Chapter 9 that allometric scaling relationships could only be obtained when considering cities that belong to the same system of cities. The fact that we observe scalings when taking a single country into account, and a cloud of points when mixing two different countries, is a signature of the integration of cities into systems of cities. It is not clear at the moment what mechanisms are responsible for the coherence that permits the existence of scaling at the system level. But clearly, the fact that cities are tightly connected through the flow of commodities, populations, information and funds must be a key factor.

Now, the same connections may be responsible for the scaling relationships themselves, and the value of the exponent. As an example, Pumain et al. [187] study the scaling of the number of employees from different economic sectors in France with population size (see also Chapter 8). They find that the number of employees in innovative sectors (such

as research and development) scales superlinearly with population size, while the number of employees in mature economic sector (such as the manufacture of food products) scales sublinearly with population size. Using historical data, they further show that the scaling behaviour of some activities has significantly changed over time: the exponent of manufacturing activities has continuously decreased since 1960, while that of research and development has continuously increased. This could be explained, they claim, by the hierarchical diffusion of innovations in systems of cities. Innovative activities first appear in large cities, entailing a larger proportion of the active population working in these sectors than in smaller cities, thus a superlinear scaling. Over time, the innovations progressively diffuse through the system of cities, the proportions are equilibrated and the value of the scaling exponent decreases.

Although the mechanism is plausible, the current issue with this interpretation is the lack of predictive model that explains the values of the various exponents.

### 10.1.3 *Cities, or systems of cities?*

So, are scaling relationships properties of cities, or of systems of cities? Probably both. The above discussion is very general, and the origin of scalings should be evaluated on a case-by-case basis. The scaling of some quantities, such as the total quantity of CO<sub>2</sub> emitted or the total length of roads are undoubtedly due to intra-urban processes (at least as long as the explanation presented in Chapter 9 holds). Indeed, the total length of roads in Los Angeles only depends on what happens in Los Angeles. Others, such as the linear scaling of total income, are probably due to the interactions of cities within the same system of cities. However, it is impossible to discriminate between both interpretations on a purely empirical basis. Ultimately, we need models that are able to reproduce at least the qualitative scaling behaviour. Plausible narratives are not enough.

## 10.2 WHAT CITIES?

As we have argued up to this point, scaling relations are a signature of various processes governing the phenomenon under study, especially when the exponent  $\beta$  is not what is naively expected [21]. However, as more and more scaling relationships are being reported in the literature, it becomes less and less clear what we really learn from these empirical findings. Mechanistic insights about these scalings are usually nonexistent, often leading to misguided interpretations.

A striking example of the fallacies which hinder the interpretation and application of scaling is given by different studies on CO<sub>2</sub> emissions due to transportation [88, 99, 171, 197]. The topic is particularly timely: pollution peaks occur in large cities worldwide with a seemingly

increasing frequency, and are suspected to be the source of serious health problems [32]. Glaeser and Kahn [99], Rybski et al [197], Fragkias et al [88], and Oliveira et al [171] are interested in how  $\text{CO}_2$  emissions scale with the population size of cities. The question they ask is simple: Are larger cities greener—in the sense that there are fewer emissions per capita for larger cities—or smoggier? Surprisingly, these different studies reach contradictory conclusions. We identify here two main sources of error which originate in the lack of understanding of the mechanisms governing the phenomenon.

The first error concerns the estimation of the quantity  $Q_{\text{CO}_2}$  of  $\text{CO}_2$  emissions due to transportation. In the absence of direct measures, Glaeser and Kahn [99] have chosen to use estimations of  $Q_{\text{CO}_2}$  based on the total distance traveled by commuters. This is in fact incorrect, and in heavily congested urban areas the relevant quantity is the total time spent in traffic [3]. Using distance leads to a serious underestimation of  $\text{CO}_2$  emissions: the effects of congestion are indeed strongly nonlinear, and the time spent in traffic jams is not proportional to the traveled distance. As a matter of fact, commuting distance and time scale differently with population size, and the time spent commuting and  $\text{CO}_2$  emissions scale with the same exponent [3].

The second, subtler, issue lies in the definition of the city itself, and over which geographical area the quantities  $Q_{\text{CO}_2}$  and  $P$  should be aggregated. There is currently great confusion in the literature about how cities should be defined, and scientists, let alone the various statistical agencies in the world, have not yet reached a consensus. For instance, the US Census Bureau defines two types of cities for statistical purposes (see Figure 29 for an illustration on the city of Minneapolis). First, the Urban Areas are defined as a set of contiguous high-density areal units with a threshold on the total population (morphological definition). The Metropolitan Statistical Areas, on the other hand, include core Urban Areas, and the areal units that sends more than a given percentage of its working population to work in the core (functional definition).

This is a crucial issue as scaling exponents are very sensitive to the way city boundaries are delineated [16].  $\text{CO}_2$  emissions are no exception: aggregating over Urban Areas or Metropolitan Statistical Areas entails radically different behaviours (see Figure 30). For the US, using the definition of urban areas provided by the Census Bureau (<http://www.census.org>), one finds that  $\text{CO}_2$  emissions per capita sharply increase with population size, implying that larger cities are less green. Using the definition of metropolitan statistical areas, also provided by the Census Bureau, one finds that  $\text{CO}_2$  emissions per capita decrease slightly with population size, implying that larger cities are greener.

Faced with these two opposite results, what should one conclude? Our point is that, in the absence of a convincing model that accounts

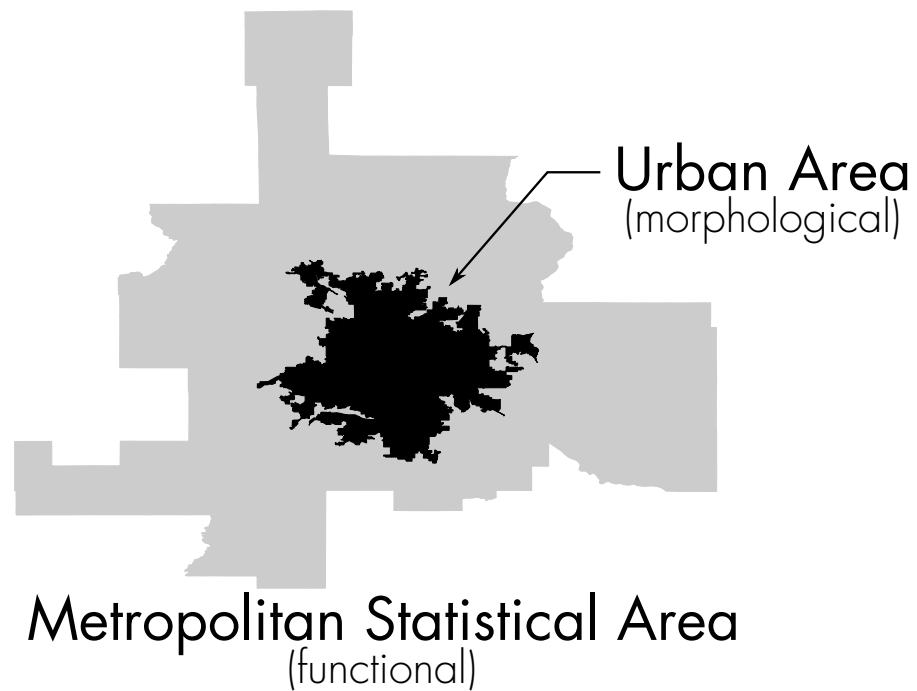


Figure 29: **City definitions in the US.** The Minneapolis Urban Area (in black) is defined by the Census Bureau as contiguous block groups with at least 1000 inhabitants per square mile. The Minneapolis-St. Paul Metropolitan Statistical Area (in grey) is defined as the counties containing the urban area as well as any adjacent county that have a high degree of integration with the core, as measured with commuting flows.

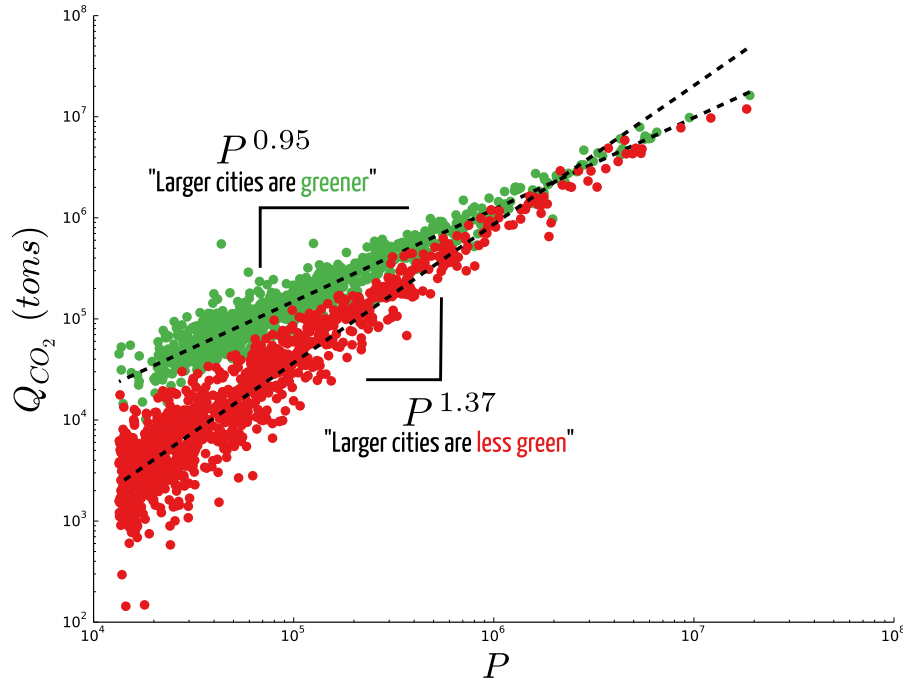


Figure 30: **Are larger cities greener or smoggier?** Scaling of transport-related  $\text{CO}_2$  emissions with the population size for US cities from the same dataset but at different aggregation levels. In red, the aggregation is done at the level of urban areas and in green for combined statistical areas. Depending on the definition of the city, the scaling exponents are qualitatively different, leading to two opposite conclusions. Data on  $\text{CO}_2$  emissions were obtained from the Vulcan Project (<http://vulcan.project.asu.edu>) (see [88, 171]). Data on the population of urban areas and metropolitan statistical areas were obtained from the Census Bureau (<http://www.census.org>).

for these differences and how they arise, nothing. Scaling relationships, and more generally data analysis, have an important role to play in the rising new science of cities. But, as the previous discussion illustrates (as well as the discussion in Chapter 4), it is dangerous to interpret empirical results without any mechanistic insight. Conclusions cannot safely be drawn from data analysis alone.

Does it mean that we should throw away scaling relationships altogether, as suggested by Arcaute et al. [16]? No, this would be tackling the problem from the wrong end. Scaling relationships are the signature of processes occurring at the system (city or system of cities) level. The issue encountered here is that the system we study is not properly defined. We don't really know what cities we are talking about!

Cities are doubtlessly a real pattern. Yet, the way we unveil this pattern with empirical data is, at best, imprecise. It is not based on a theoretical understanding of what cities are. As a result, we cannot

fully make sense of the exponents found in empirical data. We therefore believe that future research in this area should focus on

- Understanding the basic object we are working on, cities. How they should be defined, on what theoretical grounds.
- Accounting for the different qualitative behaviours of scaling exponents when different definitions are used.

Indeed, as long as we do not *know* what system we should be probing, it is not quite clear what our results mean. As long as we do not *understand* why values of exponents are different when the city definition changes, we cannot draw reasonable conclusions.

The last years have seen many scholars coming forward with policy advice based on empirical scaling relationships. It should now be clear that, given the current state of knowledge, it is a risky game. Indeed, let us consider the above CO<sub>2</sub> example: what should one do to curb CO<sub>2</sub> emissions? Favour the growth of large urban areas or the repartition of population in less populated cities? Both can be argued by considering data analysis alone. It should therefore be obvious that, until they have a satisfactory understanding of the mechanisms responsible for the observed behaviours, scientists should refrain from giving policy advice that might have unforeseen, disastrous consequences. If they choose to do so anyway, policy makers should be wary about what is, at best, a shot in the dark

### 10.3 CONCLUSION AND PERSPECTIVE

Scaling laws are useful tool to probe the internals of cities, but they are not everything. They provide an extraordinarily easy way to explore the properties of urban systems: the amount of data required is minimal, the statistical treatment trivial. Allometric scaling is thus useful to declutter the field of investigation, help clear a couple of paths, and establish a large-scale understanding of the system. But this is done at the expense of an extensive coverage of the underlying phenomena. Scalings can be seen as a gateway to the study of cities, but they cannot be the study itself.

Furthermore, there are pressing issues that need to be solved if we want to make sense of these empirical results. First, we need to question the definition of cities, and understand what systems exactly we are studying. Second, measuring exponents is not enough, and we need to understand the main processes that are responsible for the measured values. This is what we have tried to do in the previous chapter.



## Part IV

### SEGREGATION

Residential segregation is a reality. A reality so rife that it has pervaded even our everyday language through the expressions 'poor neighbourhood' or 'rich neighbourhood'. But despite its intuitive appeal, segregation is difficult to define. In this part, we propose to define segregation as a deviation to the unsegregated city, thereby providing a firm theoretical basis for any study of segregation patterns. We further propose a measure of attraction/repulsion of the different categories, which allows us to define unambiguously income classes from the original categories. We also study the properties of neighbourhoods in which the different classes concentrate, and revisit the traditional poor center/rich suburb dichotomy.



## WHAT SEGREGATION IS NOT

*The limits of my language  
Mean the limits of my world.*

— Ludwig Wittgenstein [228]

## 11.1 STUDYING SEGREGATION

We cannot judge the spatial repartition of people. There is no criterion of ‘good’ or ‘bad’ for the way people arrange themselves, no moral values attached to any spatial pattern. It is the *processes* that lead to such patterns, the intentions behind people’s decisions that make segregation condemnable. It is the *consequences* of segregation that may make it undesirable, something worth fighting against.

As a matter of fact, social residential segregation has terrible consequences. As shown in [150], residential segregation is the cause of major economic disadvantages that affect the least affluent segments of the population, through the isolation from social networks, or the presence of deficient public service in the poorest areas. Worse, it has been shown that increased levels of segregation in urban areas is associated with a higher mortality burden [140]. For all these reasons, there is a somewhat urgent need to measure the extent of segregation, especially its local component, and understand the underlying mechanisms.

In the literature, authors systematically design a single index of segregation for territories that can be very large, up to thousands of square kilometers [15]. In order to mitigate segregation, a more local, spatial information is however needed: local authorities need to locate where the poorest and richest concentrate if they want to design efficient policies to curb, or compensate for, the existing segregation. In other words, we need to provide a clear *spatial* information on the pattern of segregation. We need to identify the areas where levels of segregation are high.

Besides, if we want to design policy or incentives to reduce socio-spatial stratification and its consequences, we need to understand the processes at play. We need to understand why segregation patterns exist, and why they persist. Without mechanistic insights, attempts at regulating segregation may have unforeseen, possibly damaging consequences. The processes behind segregation are however unclear. Schelling’s cellular automata model [202], although intellectually stimulating, is very

limited in terms of predictions. More sophisticated models appeared recently [50, 100, 97], yet the link with the empirical reality is too thin, and processes are yet to be validated.

In fact, we believe that the lack of an appropriate model is likely due to the lack of identification of a clear structure, or clear behaviours in the data. In order to identify the processes at play, we urgently need to properly describe the spatial patterns of segregation; the dynamics of households (how they move, how their characteristics evolve over time) and neighbourhoods (how their population changes).

In the following, we will therefore focus on the empirical characterisation of the patterns of segregation. But first, we need to define what we mean when we talk about residential segregation.

### 11.2 THINK FIRST, MEASURE LATER

As stated many times, and at different periods in the sociology literature [72, 115, 149, 189], the study of segregation is cursed by its intuitive appeal. Pretty much everyone has heard of segregation, and has an opinion about it. This familiarity with the concept favours what Duncan and Duncan [72] called ‘naïve operationalism’: the tendency to force a sociological interpretation on measures that are at odds with the conceptual understanding of segregation. In their own words

[Segregation] is a concept rich in theoretical suggestiveness and of unquestionable heuristic value. Clearly we would not wish to sacrifice the capital of theoretization and observation already invested in the concept. Yet this is what is involved in the solution offered by naïve operationalism, in more or less arbitrary matching some convenient numerical procedure with the verbal concept of segregation... (Duncan and Duncan, 1955 [72])

For all its intuitive appeal, segregation is however an intricate, compound notion whose complexity only reveals itself through careful study. However tempting it is to start writing measures of segregation that seem ‘reasonable’, it is necessary to stop and think about the meaning of the notion first. We need to *think* segregation to be able to provide *useful* measures of segregation.

### 11.3 THE DIMENSIONS OF SEGREGATION

Segregation has been extensively studied in the Sociology and Geography literature. The most important conceptual heritage of this literature is the distinction between residential segregation’s different dimensions. Massey [149] first proposed a list of 5 dimensions (and related existing measures), which was recently reduced to 4 by Reardon [190].

**EVENNESS** (and clustering in the continuous limit, as shown by Reardon [190]) is the extent to which populations are evenly spread in the metropolitan area. Measures of evenness are affected by the fact that individuals are not spread uniformly across space in urban areas, disregarding of their respective category;

**EXPOSURE** is the extent to which different populations share the same residential area. This presupposes defining what is meant by ‘residential area’;

**CONCENTRATION** is the extent to which populations concentrate in their residential area;

**CENTRALISATION** is the extent to which populations concentrate in the center of the city. As we have seen in Chapter 5, the notion of center is meaningless in large, polycentric urban areas;

We will discuss in details the shortcomings of the measures currently proposed for each of these dimensions in Chapter 12.

#### 11.4 THE UNSEGREGATED CITY

The fundamental issue with the picture given by these 4 dimensions lies in the lack of a general theoretical framework in which all existing measures can be interpreted. Instead, we have a patchwork of seemingly unrelated measures that are labelled with either of the aforementioned dimensions. Already in 1986, Michael White [226] regretted the fact that segregation was never defined in the literature, and always considered as a given. Each index implied a different definition of segregation, which lead to endless debates about the virtues of such or such measure (dubbed the ‘index war’). Unknowingly, authors were trying to squeeze the social reality into existing measures. When, in fact, one should start by defining the social reality, before attempting to capture it with appropriate measures. As of today, no such definition of segregation exists. We shall begin our study of segregation patterns by an attempt at defining segregation. All the measure we propose then naturally follow.

Segregation manifests itself in different ways, which makes it very difficult to define. It is however easy to define what is *not* segregation: a spatial distribution of different categories that is undistinguishable from a uniform random situation [114]. Therefore, we propose to define segregation as the following

Segregation is any pattern in the spatial distribution of populations that significantly deviates from a situation where individuals would have chosen their residence at random (densities and overall category proportions being equal).

It is then easy to understand the different dimensions of [149, 190]: each of the dimensions correspond to a different ways in which a multi-dimensional pattern can deviate from its randomized counterpart. Our definition is perfectly agnostic with regards to the features of the population density pattern. It is also not concerned with the overall inequality levels.

In the context of residential segregation in urban areas, a natural null model is therefore the *unsegregated city*. In the unsegregated city, all households are distributed at random within the urban space with the further constraints that

- The total number  $N_\alpha$  of people belonging to a category  $\alpha$  is fixed and equal to that found in the data;
- The total number  $n(t)$  of households living in the areal unit  $t$  is fixed and equal to that found in the data.

which also fixes the total number of individuals  $N$  in the city. The problem of finding the numbers  $(n_\alpha(1), \dots, n_\alpha(T))$  of individuals belonging to a certain category  $\alpha$  in the  $T$  areal units of an unsegregated city is reminiscent of the traditionnal occupancy problem in combinatorics [83]. Their distribution is given by the multinomial distribution  $f(n_\alpha(1), \dots, n_\alpha(T))$ , and the number of people of category  $\alpha$  in the areal unit  $t$  by a binomial distribution. Therefore, in an unsegregated city, we have

$$\begin{aligned} \mathbb{E}[n_\alpha(t)] &= N_\alpha \frac{n(t)}{N} \\ \text{Var}[n_\alpha(t)] &= N_\alpha \frac{n(t)}{N} \left(1 - \frac{n(t)}{N}\right) \end{aligned} \tag{84}$$

where  $N$  is the total number of households in the city. In metropolitan areas  $N_\alpha$  is larged compared to 1, and the distribution of the  $n_\alpha(t)$  can be approximated by a Gaussian with the same mean and variance.

Most studies exploring the question of spatial segregation define measures before comparing their value for different cities. Knowing that two quantities are different is however not enough: we also have to know whether this difference is *significant*. In order to assess the significance of a result, we have to compare it to what is obtained for a reasonable null model. As we will see in Chapter 12, the unsegregated city model allows us to assess whether a given pattern is the result of a segregation process or not.

Any spatial distribution patterns could theoretically obtained via a random repartition of households. They are however not equally likely. We propose to measure the total segregation by the likelihood of obtaining a given pattern, assuming a random distribution.

In this chapter, we have discussed some of the improvements that could be brought to the existing measures in the literature. In particular, we have emphasized the need for a *local* knowledge of the patterns of segregation. We have also laid the theoretical foundation upon which we are going to design new measures. In Chapter 12, we start from the above-defined null model to propose a way to quantify the presence of various categories in parts of the city. This allows us to identify and delineate neighbourhoods, measure the interactions between the categories, and extract a class structure from the spatial pattern alone.





---

*To understand is to perceive patterns.*

— Isaiah Berlin [31]

## 12.1 INTRODUCTION

### 12.1.1 *Shortcomings in the current empirical picture*

There are many different ways in which a spatial pattern can deviate from its randomised counterpart, and at least as many different measures one could perform. In this chapter, we will try to quantify these patterns in a way that may allow us to *understand* the phenomenon of segregation.

Of course, segregation has been extensively studied in the literature. However, we identify several difficulties in the current empirical picture.

First, some issues are tied to the existence of several categories in the underlying data. Historically, measurements of racial segregation were limited to measures between 2 population groups. However, most measures generalise poorly to a situation with many groups, and the others do not necessarily have a clear interpretation [189]. Worse, in the case of groups based on a continuum (such as income), the thresholds chosen to define classes are usually arbitrary [116]. We propose to solve this issue by defining classes in a unambiguous and non-arbitrary way through their pattern of spatial interaction. Applied to the distribution of income categories in US cities, we find 3 emergent categories, which are naturally interpreted as the lower-, middle- and higher-income classes.

Second, most authors systematically design a single index of segregation for territories that can be very large, up to thousands of square kilometers [15]. In order to mitigate segregation, a more local, spatial information is however needed: local authorities need to locate where the poorest and richest concentrate if they want to design efficient policies to curb, or compensate for, the existing segregation. Furthermore, a local description of the repartition of the different categories is the first step towards the exploration of the mechanisms responsible for segregation: it is necessary to gather hints (as well as empirical regularities) that are essential to build a reasonable model. In other words, we need to provide a clear *spatial* information on the pattern of segregation.

The lack of clear spatial characterization of the distribution of individuals is not tied to the problem of segregation in particular, but pertains to the field of spatial statistics [193]. Many studies avoided this spatial problem by considering cities as monocentric and circular, and rely on either an arbitrary definition of the city center boundaries, or on indices computed as a function to the distance to the center (whatever this center may be, see Part ii). However, most if not all cities are anisotropic, and the large ones, polycentric (see Chapter 5), casting some doubt about the application of the monocentric city picture. Many empirical studies and models in economics aim to explain the difference between central cities and suburbs [100, 50]. Yet, the sole stylized fact upon which they rely – city centers tend to be allegedly poorer than suburbs (in the US) – lacks a solid empirical basis.

In the following, we propose to answer the following questions

- How can we quantify the presence of the different categories in areal units? Can we say whether they are overrepresented or normally represented? How can we define neighbourhoods?
- Can we quantify interactions between the different categories?
- Can we define meaningful classes from the original data?
- Do classes tend to leave in geographically coherent areas, or are they scattered across the city?
- Is there a difference between the city center and the suburbs? How can we quantify this adequately?

### 12.1.2 Notations

In the following, we will illustrate our measures using data from the 2000 US Census on the income of households per Census blockgroup. Data present themselves as a number of households per blockgroup, sorted in different income categories. There are  $N$  individuals and  $T$  tracts in the considered geographical area, and we note  $N_\alpha$  the number of individuals belonging to the category  $\alpha$ . Finally, we write  $n(t)$  the total number of individuals living in the tract  $t$ , and  $n_\alpha(t)$  the total number of individuals who belong to category  $\alpha$  living in the tract  $t$ .

## 12.2 PRESENCE OF CATEGORIES

In order to quantify segregation, we first need to measure the extent to which categories are spread unevenly across space. Therefore, we start our analysis with a discussion on how to quantify the presence of a category in areal units. Several indicators exist, and one needs to be aware of their meaning, their qualities and their shortcomings.

### 12.2.1 Concentration index

The concentration index measures the proportion of individuals from category  $\alpha$  in the areal unit  $t$ .

$$c(t) = \frac{n_\alpha(t)}{N_\alpha} \quad (85)$$

The concentration is composition-invariant: it does not depend on the relative proportion of category  $\alpha$  in the geographical zone as a whole.

Nevertheless, its value strongly depends on the total population of the areal unit we are studying: more populated areal units mechanically entail higher values of concentration. Segregation measures based on the concentration (such as the dissimilarity index) will therefore be dominated by the values in highly populated areal units. This also makes values of concentration difficult to interpret: we don't know whether large (respectively low) values of concentration are the result of a large (respective low) population, or of a local concentration of individuals in the area.

### 12.2.2 Proportion index

Sometimes, we would prefer to know the proportion of individuals of a given category in a unit. In our notations, the proportion index is simply defined as

$$p(t) = \frac{n_\alpha(t)}{n(t)} \quad (86)$$

Although the values of the proportion index are easier to interpret ("x% of the individuals living in this areal unit belong to such category"), they are not a good indicator of segregation.

Indeed, they strongly depend on the relative proportion of individuals of the category in the geographical area being studied. For instance, in a city where 90% of the individuals belong to category  $A$ , the proportion of people belonging to category  $A$  is very likely to be high in all areal units in the city. The measure of proportion is therefore strongly tied to the overall inequality levels.

### 12.2.3 An unbiased measure: representation

#### 12.2.3.1 Definition

The representation solves the problems linked to both measures of concentration and proportion. The idea behind the measure of representation is that segregation is, as we argued in Chapter 11, a departure from the situation where households would be spatially distributed at

random. The properties of such a ‘random’, unsegregated city are well known, and the distribution of categories in each areal unit is given by a binomial distribution. The representation is thus defined as the number  $n_\alpha(t)$  divided by its expected value in an unsegregated city,  $N_\alpha \frac{n(t)}{N}$

$$r_\alpha(t) = \frac{n_\alpha(t)/n(t)}{N_\alpha/N} \quad (87)$$

Another way to understand the representation is to compare it to the above-defined concentration and proportion. We can indeed write

$$r_\alpha(t) = \frac{c(t)}{n(t)/N} = \frac{p(t)}{N_\alpha/N} \quad (88)$$

The representation can thus be interpreted as the concentration normalised by the local population concentration, or the proportion renormalised by the proportion of the category at the city level, thereby addressing the aforementioned shortcomings.

### 12.2.3.2 Measuring significant deviations

The representation  $r_\alpha(t)$  takes values between 0 (when no individuals from the category  $\alpha$  are present in  $t$ ) and  $\frac{N}{N_\alpha}$  (when all individuals in  $t$  belong to the category  $\alpha$ ). In a city where individuals are distributed uniformly (see Chapter 11),  $r_\alpha(t) = 1$  in every tract  $t$ .

In an unsegregated situation, the values of the representation are likely to be close to 1, but not necessarily strictly equal to 1. There is indeed a non-zero probability for any distribution to be obtained by chance. It is therefore not obvious whether a given value of representation could have been obtained in the unsegregated configuration. However, to quantify segregation, we need to know how *likely* it is that the present pattern is not the result of a random repartition of individuals. In other words, we need to know whether areal units depart *significantly* from the unsegregated situation.

The distribution of individuals in a tract  $t$  in the unsegregated city follows a binomial distribution. We can therefore easily compute how likely it is that the representation  $r_\alpha(t)$  we measure has been obtained by chance. To do that, we first compute the variance of the representation in the unsegregated configuration:

$$\text{Var}[r_\alpha(t)] = \sigma_\alpha(t)^2 = \frac{1}{N_\alpha} \left[ \frac{N}{n(t)} - 1 \right] \quad (89)$$

We say that the representation departs *significantly* from the unsegregated configuration if we can be sure with 99% confidence that the pattern has not been obtained at random. It follows that

- $\alpha$  is **overrepresented** in  $t$  iff  $r_\alpha(t) > 1 + 2.57 \sigma_\alpha(t)$
- $\alpha$  is **underrepresented** in  $t$  iff  $r_\alpha(t) < 1 - 2.57 \sigma_\alpha(t)$

Note that the expression of the representation (Eq. 87) is very similar to the formula used in economics to compute comparative advantages [19], or to the localisation quotient used in various contexts [15, 204]. To our knowledge, however, this formula has never been justified by a null model in the context of residential location.

The representation allows to assess the significance of the deviation of population distributions from the unsegregated city. As we will show below, it is also the building block for measuring the level of repulsion or attraction between populations – allowing us to uncover the different classes – and to identify the neighbourhoods where the different categories concentrate.

Last, but not least, the representation defined here does not depend on the class structure at the city scale, but only on the spatial repartition of individuals belonging to each class. This is essential to be able to compare different cities where the group compositions – or inequality – might differ. Inequality and segregation are indeed two separate concepts, and the way they are measured should be distinct from one another. In that sense, the representation is preferable to the measures of concentration or representation as a basis to quantify segregation.

## 12.3 MEASURING THE ATTRACTION AND REPULSION OF CATEGORIES

### 12.3.1 *Exposure*

If we want to uncover the mechanisms underlying segregated patterns, it is important to measure and understand the interactions between categories. However, existing measures do not allow to quantify to which extent different populations attract or repel one another. What we mean here by interaction is the co-presence of the different categories in the same areal units, thus potential interactions. This is the best one can do in the absence of data on the actual interactions between individuals.

The measure we define is inspired by the M-value first introduced by Marcon & Puech in the economics literature [146] and used as a measure of interaction in [118]. These authors were interested in measuring the geographic concentration of different types of industries. While previous measures (such as Ripley's K-value) allow to identify departures from a random (Poisson) distribution, the M-value's interest resides in the possibility to evaluate different industries' tendency to co-locate.

The idea, in the context of segregation, is simple. We consider two categories  $\alpha$  and  $\beta$  and we would like to measure to which extent they are co-located in the same areal unit. Essentially, we measure the repre-

sensation of the category  $\beta$  as witnessed on average by the individuals in category  $\alpha$ , and obtain the following quantity  $E_{\alpha\beta}$

$$E_{\alpha\beta} = \frac{1}{N_\alpha} \sum_{t=1}^T n_\alpha(t) r_\beta(t) \quad (90)$$

Although it is not obvious with this formulation, this measure is symmetric:  $E_{\alpha\beta} = E_{\beta\alpha}$ . Effectively, the E-value is a measure of exposure, according to the typology of segregation measures found in [149]. It is however different from the traditional measure of exposure found in the literature [30], as it allows to distinguish between the situations where categories attract, or repel one another.

In the case of an unsegregated city, every household in  $\alpha$  sees on average  $r_\beta = 1$  and we have  $E_{\alpha\beta} = 1$ . If populations  $\alpha$  and  $\beta$  *attract* one another, that is if they tend to be overrepresented in the same areal units, every household  $\alpha$  sees  $r_\beta > 1$  and we have  $E_{\alpha\beta} > 1$  at the city scale. On the other hand, if they *repel* one another, every household  $\alpha$  sees  $r_\beta < 1$  and we have  $E_{\alpha\beta} < 1$  at the city scale.

### 12.3.2 Extreme values

The minimum of the exposure for two classes  $\alpha$  and  $\beta$  is obtained when these two categories are never present together in the same areal unit. Then

$$E_{\alpha\beta}^{\min} = 0 \quad (91)$$

and the theoretical maximum is obtained when the two classes are alone in the system and otherwise distributed at random

$$E_{\alpha\beta}^{\max} = \frac{N^2}{4 N_\alpha N_\beta} \quad (92)$$

These extrema are useful when comparing the exposure values for different categories, and across different cities.

### 12.3.3 Isolation

In the case  $\alpha = \beta$ , the previous measure represents the ‘isolation’ defined as

$$I_\alpha = \frac{1}{N_\alpha} \sum_{t=1}^t n_\alpha(t) r_\alpha(t) \quad (93)$$

and measures to which extent individuals from the same category are exposed to their kins. In the unsegregated city, where individuals are indifferent to others when choosing their residence, we have  $I_{\alpha}^{min} = 1$ . On the other hand, in the extreme situation where individuals belonging to the class  $\alpha$  live isolated from the others, the isolation reaches its maximum value

$$I_{\alpha}^{max} = \frac{N}{N_{\alpha}} \quad (94)$$

## 12.4 EMERGENT SOCIAL CLASSES

### 12.4.1 *Defining classes*

The study of income segregation must be rooted in a particular definition of categories (or classes). There is however no consensus in the literature about how to separate households in different classes according to their income, and studies generally rely on more or less arbitrary divisions. While in some particular cases grouping the original categories in pre-defined classes is justified, most authors do so for mere convenience. However, as some sociologists have already pointed out [80], imposing the existence of absolute, artificial entities is necessarily going to skew our reading of the data. Entities such as social classes do not have an existence of their own. Grouping the individuals into arbitrary classes when studying segregation is thus problematic: it amounts to imposing a class structure on the society before assessing the existence of this structure (which manifests itself by the differentiated spatial repartition of individuals with different income, segregation). Furthermore, in the absence of recognized standards, different authors will likely have different definitions of classes, making the comparisons between different results in the literature difficult.

Here, instead of imposing an arbitrary class structure, we let the class structure emerge from the data themselves. Our starting hypothesis is the following: if there is such a thing as a social stratification based on income, it should be reflected in the households' behaviours. The hypothesis is that households belonging to the same class should tend to live together, while households belonging to different classes should tend to avoid one another (It is worth noting that this *horizontal* definition of segregation is not relevant in every context; in the 19th century Paris for instance, segregation was also vertical, with rich families living in the lowest floors of buildings while poor individuals did tend to live in the highest flats). The idea is thus to define classes based on the way they manifest themselves through the spatial repartition of the different categories. Of course, spatial proximity does not necessarily imply social proximity. In particular, Chamboredon showed that in some

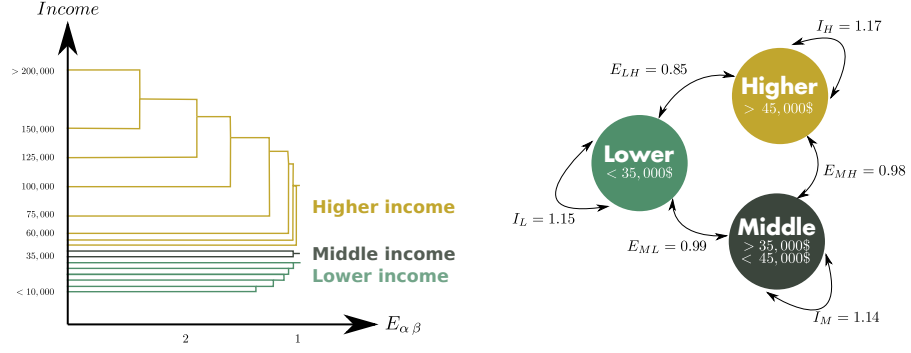


Figure 31: **Emergent classes.** (Left) Alluvial diagram showing the successive aggregations of income categories in the clustering process, and the value of the exposure at which the aggregation took place. The aggregation stops when there is no pair of category for which  $E > 1$ , that is when all classes are at best indifferent to one another. One can see on this diagram that the highest income categories attract one another more (higher values of  $E_{\alpha\beta}$ ) than the lowest income categories. (Right) The classes that emerge from our analysis, and their respective exposure and isolation values. The lower and higher income classes repel one another, while the middle income class is indifferent to either other classes. The higher-income class is slightly more coherent than the lower-income, which is more coherent than the middle-income class, as reflected by the isolation coefficient  $I$ .

*The work of  
Chamboredon was  
kindly brought to  
my attention by  
Yann Renisio.*

big French housing projects, households belonging to different social classes were artificially brought in close proximity to one another but did not necessarily interact with one another [54]. We thus assume here that the social class of housing tenants is not determined in a top-down fashion, so that the spatial repartition of different income classes reflects the nature of the interaction between these classes.

#### 12.4.2 Income classes in the US

We choose as a starting point the finest income subdivision given by the Census Bureau (16 subdivisions) and compute the  $16 \times 16$  matrix of  $E_{\alpha\beta}$  values for all cities. We then perform a hierarchical clustering on this matrix, successively aggregating the subdivisions with the highest  $E_{\alpha\beta}$  values. We stop the aggregation process when the only classes left are indifferent ( $E_{\alpha\beta} = 1$  with 99% confidence) or repel one another ( $E_{\alpha\beta} < 1$  with 99% confidence) [7]. We obtain the dendrogram presented on Figure 31.

Strikingly, the outcome of this method is the emergence of 3 distinct classes: the higher-income (47% of the US population) and the lower-income (42% of the US population) classes – which repel one another strongly while being respectively very coherent – and a somewhat meagre middle-income class (11% of the population) that is relatively indifferent to the other classes. This result implies that there is some



truth in the conventional way of dividing populations into 3 income classes, and that what we casually perceive as the social stratification in our cities actually emerges from the spatial interaction of people. Surprisingly, however, the middle-income class as obtained here represents a significantly smaller part of the population than other definitions.

Our method has several advantages over a casual, arbitrary definition: it only depends on single tunable parameter, the size of the confidence interval. Although, once an agreement has been reached, the class structure does not depend on who is performing the analysis. Its origins are tractable, and can be argued on a quantitative basis. Because it is quantitative, it allows comparison of the stratification between different points in time, or between different countries. It can also be compared to other class divisions that would be obtained using a different medium for interaction, for instance mobile phone communications [78].

In the following, we will systematically use the classes thus obtained.

## 12.5 LARGER CITIES ARE RICHER

At the scale of an entire country, segregation can manifest itself in the unequal representation of the income classes in different urban areas. We plot on Figure 32 the ratio  $N_\alpha^>(H)/N^>(H)$  where  $N^>(H)$  is the number of cities of population greater than  $H$ , and  $N_\alpha^>(H)$  the number of cities of population greater than  $H$  for which the class  $\alpha$  is overrepresented.

A decreasing curve indicates that the category  $\alpha$  tends to be underrepresented in larger urban areas, while an increasing curve shows that the category  $\alpha$  tends to be overrepresented in larger urban areas. The representation is measured with respect to the total population at the US level.

There is a clear differentiation between cities: among the 276 MSA in our dataset, no city exhibits a number of households per class that is representative of the US as a whole. Furthermore, the number of cities where higher-income households are overrepresented increases with the size of the cities, while the inverse trend is true for lower-income households. Therefore, larger cities are not richer in the sense that rich households tend to be overrepresented in large cities, and underrepresented in small ones.

Surprisingly, this effect is not visible using the Gini coefficient (see Figure 32). This hints at the limitations of the Gini index to compare income inequalities across an entire country.

## 12.6 DELINEATING NEIGHBOURHOODS

### 12.6.1 *Defining neighbourhoods*

Now that we can identify the areal units where classes are overrepresented, how can we delineate neighbourhoods?

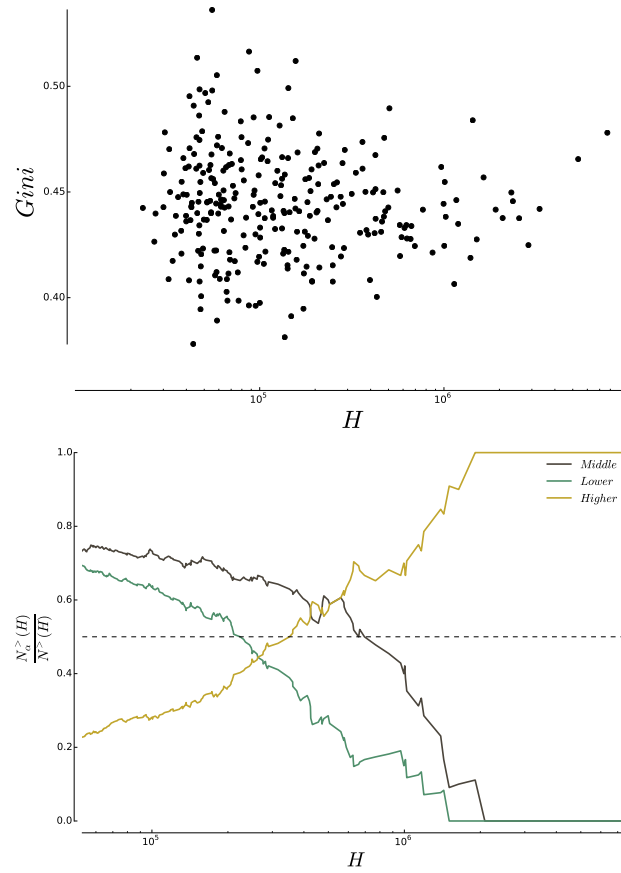


Figure 32: **Larger cities are richer.** (Top) Gini coefficient of the income distribution of the 280 MSA in 2000 versus the number of households in the city. As one can see, there is no clear trend. (Bottom) Proportion of cities in which the different classes are overrepresented, as a function of the total population of the city. One can clearly see that as cities get larger rich people will be overrepresented and poor people underrepresented (compared to national levels).

Considering a category  $\alpha$ , we first look for the areal units where the category is overrepresented. We then consider that two areal units in this set are part of the same neighbourhood if they are contiguous. Of course, this approach has limitations (some remarks that sprung in the discussion on the different methods to find activity centers in Chapter 5 are relevant in this context too), but it gives us a reasonable definition of neighbourhoods to work with. Let us now focus on the properties of these neighbourhoods.

### 12.6.2 Clustering

Intra-tract measures such as the exposure are not enough to quantify segregation. Indeed, areal units where a given class is overrepresented can arrange themselves in different ways, without the intra-tract measures of segregation being affected [225]. In order to illustrate this, we



Figure 33: **Neighbourhoods.** The neighbourhoods in Atlanta for the three different income category. In black, the tracts where the corresponding class is overrepresented, in white where it is underrepresented and in grey where its value is indistinguishable from the random distribution. All MSA defined for the 2000 Census exhibit a total exclusion between lower-income and higher-income neighbourhoods: the pictures for lower- and higher-income classes are the perfect negative of one another. In contrast, middle-income households are scattered across the city.

consider the schematic cases represented on Figure 34, and assume that they are obtained by reshuffling the various squares around. Obviously, the checkerboard on the left depicts a very different segregation situation from the divided situation on the right while intra-tract measures would give identical results.

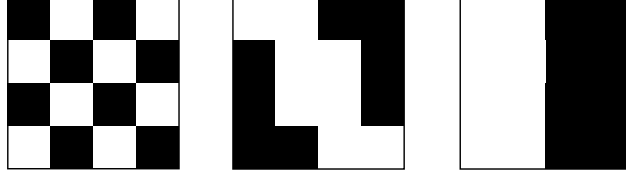


Figure 34: **Spatial considerations.** Three situations that are identical for intra-areal unit measures, but that represent different segregation levels. (Left) The checkerboard city popularised by White [225], corresponding to a clustering value (defined in Eq. 95) of  $C = 0$  for the black squares. (Middle) An intermediate situation between the checkerboard and the divided city, corresponding to  $C \approx 0.86$ . (Right) The divided city, corresponding to  $C = 1$ .

A way to distinguish between different spatial arrangements is then to measure how clustered the overrepresented areal units are. We first aggregate adjacent overrepresented areal units (for a given class) leading to consistent neighbourhoods. The ratio of the number  $N_n$  of neighborhoods (clusters) to the total number  $N_o$  of overrepresented areal units measures the level of clustering and in

$$C = \frac{N_o - N_n}{N_o - 1} \quad (95)$$

such that this quantity is  $C = 0$  in a checkerboard-like situation, and  $C = 1$  when all areal units form a unique neighbourhood. We show on Figure 35 the distribution of  $C$  for the three classes over all cities in our dataset. As one could infer from the maps on Figure 33, the rich and poor areal units are well clustered, with a respective average clustering of  $C = 0.80$  and  $C = 0.74$ . The Middle class is on the other hand less coherent, with a average clustering  $C = 0.55$ .

### 12.6.3 Concentration in neighbourhoods

If a given class is overrepresented in a neighbourhood, it does not however mean that most of the individuals belonging to this class live in this neighbourhood. We compute the ratio of households of each income class that lives in a neighbourhood over the total number of individuals in the income class (for rich, poor, and middle class). Results (Figure 36) indicate that essentially less than 50% of each class live in their respective neighbourhood, while the rest is dispatched over the rest of the

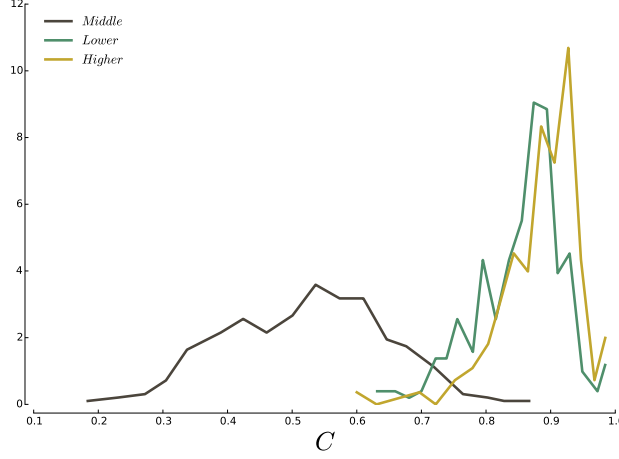


Figure 35: **Clustering coefficient.** Distribution of the value of the clustering coefficient for all cities in our dataset, for the 3 classes. The higher income class exhibits the highest level of clustering, with an average of  $\bar{C} = 0.90$ , followed by the lower income class with on average  $\bar{C} = 0.87$ . The Middle income class households are significantly less clustered than the previous two, with  $\bar{C} = 0.56$  on average.

city. The average concentration decreases from higher-income individuals (50%), to lower-income (48%) and middle-income individuals (32%).

#### 12.6.4 One large neighbourhood, or several small ones?

Finally, large values of clustering can hide different situations. We could have on one hand a ‘giant’ neighbourhood and several isolated areal units, which would essentially mean that each class concentrates in a unique neighbourhoods. Or on the other hand, several neighbourhoods of similar sizes, meaning that the different classes concentrate in several neighbourhoods across the city. In order to distinguish between the two situations, we plot

$$P = H_2^N / H_1^N \quad (96)$$

where  $H_1^N$  is the population of the largest neighbourhood, and  $H_2^N$  the population of the second largest neighbourhood. The results are shown on Figure 37, and again show a different behaviour for the middle-income on one side, and higher-income and lower-income on the other side. The size of the middle-income neighbourhoods are relatively balanced, with on average  $P = 0.62$ . Higher- and lower-income neighbourhoods, on the other hand, are dominated by one big neighbourhood, with respectively  $P = 0.22$  and  $P = 0.26$  on average.

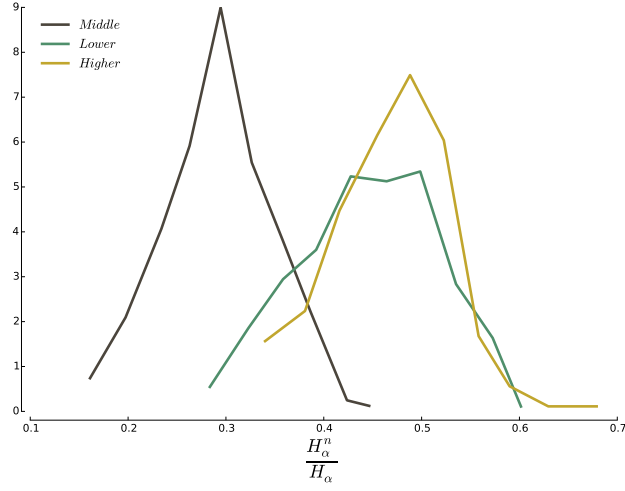


Figure 36: **Concentration in neighbourhoods.** Distribution of the fraction of households belonging to a given class and that live in a neighbourhood where it is overrepresented (Middle, Lower, or Higher).

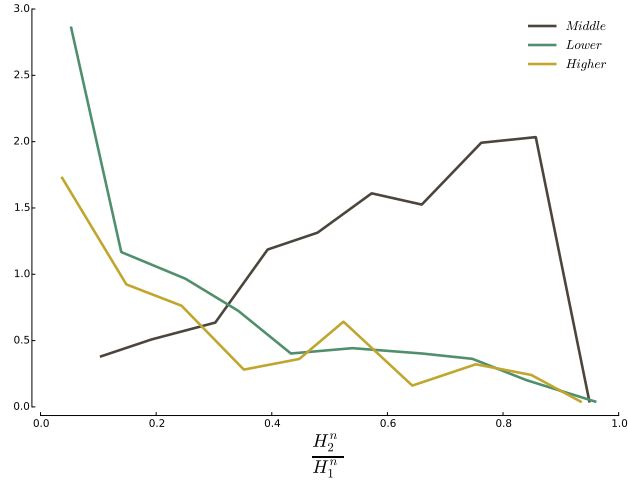


Figure 37: **Poly-neighbourhoods.** Distribution of the ratio of the size of the largest and second largest neighbourhoods for each class for all MSA in the US. Higher- and lower-income households tend to concentrate in single neighbourhood, with a secondary center that is on average 22% and 26% the size of the largest one, respectively. Middle-income households tend to be more dispersed, with a secondary neighbourhood that is on average 62% of the size of the largest.

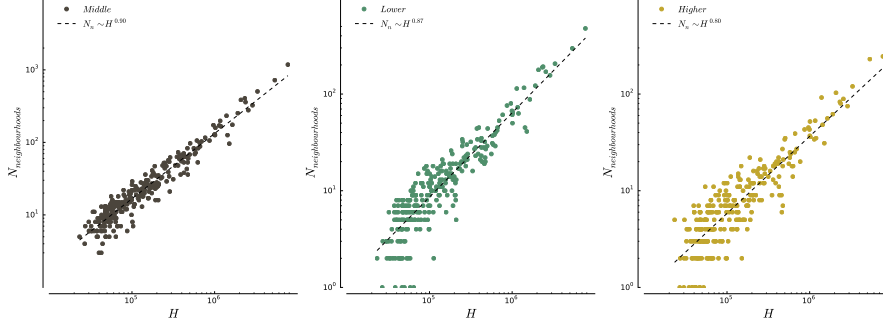


Figure 38: **Number of neighbourhoods and city size.** Number of neighbourhoods for the three different classes as a function of the size of the city. These plots in loglog show that we have a behavior consistent with a power law with exponent less than one (and with different value for each class), with  $r^2$  values that range between 0.88 (higher-income) and 0.96 (middle-income). Combined with the linear increase of the number of over-represented units with the number of households, this sub-linear increase in the number of neighbourhoods shows the tendency of classes to cluster more as cities get larger.

#### 12.6.5 *Scaling of the number of neighbourhoods*

The clustering values are high, indicating that the neighbourhoods occupied by households of different classes are very coherent. We can now wonder whether there is an effect of the city size on the number of neighbourhoods. We plot on Figure 38 the number of neighbourhoods found for all three classes as a function of population. For each class, The curve is well-fitted by a powerlaw function of the form

$$N_n = b H^\beta \quad (97)$$

where the exponent  $\beta$  is less than one and depends on the class, indicating that there are proportionally less neighbourhoods in larger cities (the number areal units scales proportionally with the population size). The values of the exponents are

$$\beta_H = 0.80$$

$$\beta_L = 0.87$$

$$\beta_M = 0.90$$

One is tempted to conclude from these numbers that the different classes become more spatially coherent as the population increases. Yet, this conclusion only holds if the number of areal units in which each class is overrepresented does not itself vary sublinearly with population size. We plot on Figure 39 these numbers as a function of the size of

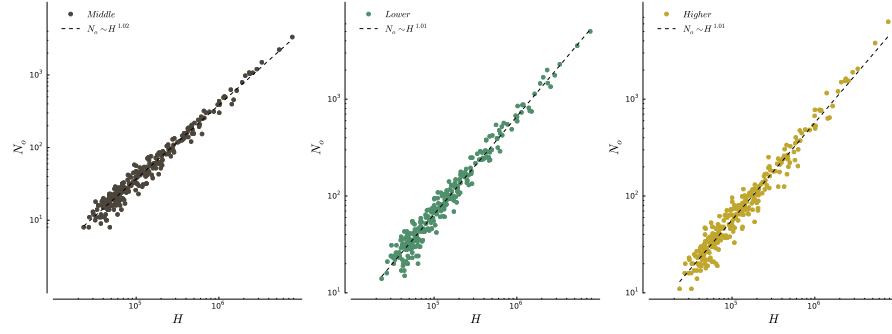


Figure 39: **Number of overrepresented areal units and city size.** Number of areal units where each class is overrepresented as a function of the total number of households in the city. The behaviour is consistent with a linear behaviour in the three cases.

the city. We find that the behaviour of the number of overrepresented units is consistent with a linear behaviour for all three classes. Together with the exponents above, this shows that the tendency of the classes to cluster is greater as the city size increases.

In other words, the different classes are more spatially isolated as the city size increases, implying higher levels of spatial segregation. We note that the phenomenon is more important for higher-income households than for lower- and middle-income households, justifying to an extent the existence of the expression ‘ghettos for the rich’.

## 12.7 POOR CENTERS, RICH SUBURBS?

In many studies, the question of the spatial pattern of segregation is limited to the study of the center versus suburb and is usually addressed in two different ways. First, a central area is defined by arbitrary boundaries and measures are performed at the scale of the so-called center and at the scale of the rest, labelled as ‘suburbs’. The issue with this approach is that the conclusions depend on the chosen boundaries and there is no unique unambiguous definition of the city center: while some consider it to be the Central Business District [100], others choose to define the center as the urban core (urbanized area), where the population density is higher. The second approach, in an attempt to get rid of arbitrary boundaries, consists in plotting indicators of wealth as a function of distance to the center [100]. This approach, inspired by the monocentric and isotropic city of many economic studies such as the Von Thünen or the Alonso-Muth-Mills model [49], has however a serious flaw: cities are not isotropic and are spread unevenly in space, leading to very irregular shapes [144]. Representing any quantity versus the distance to a center thus amounts to average over very different areas and is necessarily misleading in clear polycentric cases (as it is the case for large cities [2]. See also Chapter 5). The notion of distance



to the center is indeed meaningless in polycentric situations.

We propose here a different approach that does not require the definition of a distance to the center. Instead, we plot the average representation computed over all areal units (Census blockgroups in this dataset) with a given density population  $\rho$ , as a function of the density  $\rho$ . Indeed, what is usually meant by ‘center’ of a city are the areas with the highest residential (or employment) densities.

Our findings shed a new light on the difference of social composition between the high-density and low-density areas in cities. As shown on Figure 40, we find that rich households are overrepresented in low-density regions on average. While this agrees well with the opinion people have of suburban America, there is a more surprising result: higher income households are also overrepresented in areas with very large densities (typically above 20,000 inhabitant/km<sup>2</sup>). In between, neighborhood with intermediate values of density (between 1,000 and 20,000 inhabitants/km<sup>2</sup>), are lower-income neighbourhoods.

Only few cities in the US have neighbourhoods that reach the threshold of 20,000 inhabitants per km<sup>2</sup>, which can explain why we observe in most cases poor centers and rich suburbs. We can wonder whether the difference usually discussed between North American and European cities does not come, in fact, from differences in terms of densities.

## 12.8 CONCLUSION AND PERSPECTIVE

Instead of attempting to define segregation by enumerating its different aspects, we took a radically different – yet simpler – approach. We chose to define segregation through specifying what it is not. This naturally lead to defining the measure of representation, which is used in turn to delineate neighbourhoods. We further defined the exposure (still based on the representation), which measures the extent to which different categories attract, repel or are indifferent to one another.

We then showed that we can define classes in a non-parametric way and 3 main income classes emerged for the 2000 US Census data. The middle-income class corresponds to a smaller income range than what is usually admitted, a curiosity that certainly deserves further investigations. In terms of spatial arrangement, although the fraction of the population that is contained in neighbourhoods does not change with city size, the neighbourhoods are geographically more coherent as cities get larger, which corresponds in effect to an increased level of segregation as the size of the city increases. The behaviors of different categories are very coherent and we showed that we could simplify the description of these complex systems by reducing the sometimes large number of categories to a small number of classes. This is an important point which will simplify the description and modeling of stratification mechanisms.

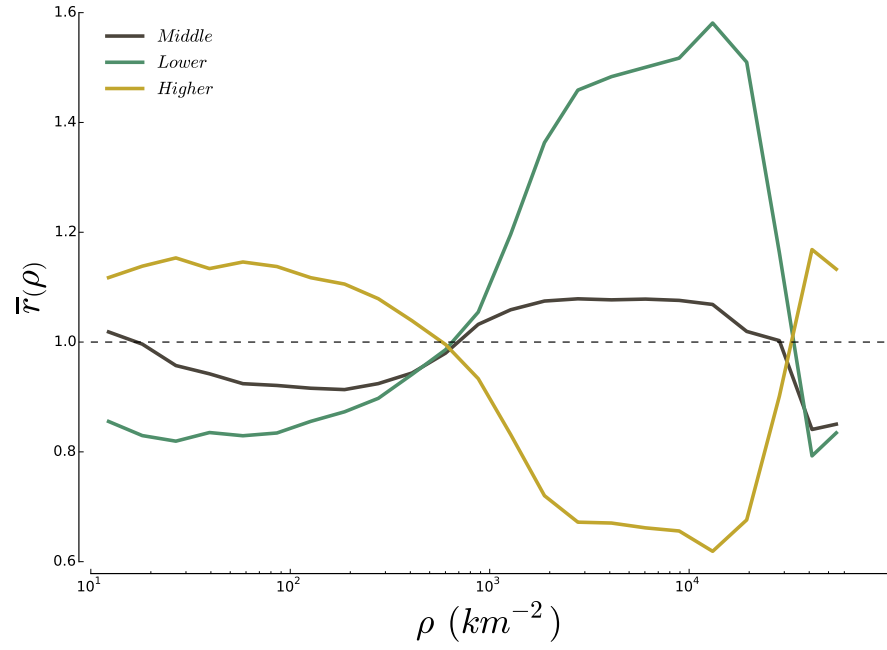


Figure 40: **Representation and density.** Average representation of the higher-, middle- and lower-income classes over the 276 MSA as a function of the local density of households. On average, we find that low-density regions (the suburbs) are rich, while high density regions (the center) are poor, confirming empirically on a large dataset a stylized facts that had previously emerged from local studies. Interestingly, we also find that very large density areas ( $\rho > 20,000/km^2$ ) are rich on average, suggesting that density may be one relevant element in an eventual explanation of the differences between neighbourhoods [113].

Our results point to the intriguing fact that higher-income households are on average overrepresented in very dense areas. Such high density areas are relatively rare in the US, which might explain in part why authors have traditionally simplified the picture, talking about poor centers and rich suburbs. This result echoes Jane Jacobs' analysis [113] that neighbourhoods with the highest dwelling densities usually are the ones exhibiting the most vitality, and therefore the most attractive. Of course, high densities are not everything, and some high-density neighbourhoods also are lower-income neighbourhoods. Further investigations along these lines may provide quantitative insights into the mechanisms leading to urban decline or urban regeneration.

In this Chapter, we have tried to highlight the *spatial* pattern of segregation. We believe that the identification of neighbourhoods that our method permits will allow a finer-scale investigation of these spatial patterns. The fundamental issue that runs beneath, however, is the need for a useful, simplified description of spatial density. A problem yet to be solved, but that has a huge potential of applications. We note that the problem is tightly linked, if not identical, to the one we encountered while trying to describe the spatial distribution of density in Chapter 7.



## Part V

### URBAN NETWORKS

People, energy, information and goods are carried through cities (and across systems of cities) thanks to various networks. In this part, we succinctly present our work on these—spatial—networks.

We first propose a quantitative method to classify cities that is based on a new perspective on street patterns, and the use of the OpenStreetMaps database. In the second chapter, we propose a model for the growth of spatial networks based on cost-benefit analysis. The resulting networks exhibit a crossover between the star graph and the minimum spanning tree when the ratio of costs and benefits evolve. In the intermediate regime, the networks adopt a hub-and-spoke hierarchical structure that has many interesting properties. We conclude this part with a large-scale description of subway and railway networks. Using the model presented in the previous chapter, we are able to predict many of their properties based on the characteristics of the underlying city or country.



## A TYPOLOGY OF STREET PATTERNS

The following chapter is a reprint of an article, *A typology of street patterns*, that was previously published by the author of this thesis with Marc Barthelemy [6].

Street networks of cities can be thought as a simplified schematic view of cities, which however captures a large part of their structure and organization [213]. Despite their apparent diversity, underlying universal mechanisms are certainly at play in the formation and evolution of street networks and extracting common patterns between cities is a way towards their identification. This program is not new [107], but the recent dramatic increase of data availability such as digitized maps, historical or contemporary [218, 24, 181] allows now to test ideas and models on large scale cross-sectional and historical data.

Streets form a network which to a good approximation is planar (where nodes are intersections and links are segment roads) and which is now fairly well characterized [120, 194, 179, 180, 129, 65, 53, 229, 119, 152, 55, 63]. Due to spatial constraints, the degree distribution is peaked, the clustering coefficient and assortativity are large, and most of the interesting information lies in the spatial distribution of betweenness centrality [22]. It is then tempting to use this information to compare various cities with each other and to provide a classification.

The problem, from a fundamental point of view is however difficult: finding a typology of street patterns amounts essentially to classify planar graphs, a non trivial problem. For street networks, this problem has been addressed by the space syntax community [110, 177] and a good account can be found in the book by Marshall [147]. These works, although based on empirical observations, contain a large part of subjectivity and our goal is to eliminate this subjective part to reach a non ambiguous, scientific classification of these patterns. An interesting direction was provided in the study of leaves and their classification according to their veination patterns [122, 160], but with a notable difference which prevents us from a direct application to streets and which is the existence of a hierarchy of veins governed by their diameter. From a mathematical point of view there exists an exact bijection between planar graphs and trees [45] which provides an interesting direction. Using this bijection, classifying planar graphs would amount to classify trees, which is a simpler problem. However, this bijection does not take into account the geometrical shape of the planar graph: indeed two street patterns can have the same topology but cells could be of very

OpenStreetMap  
data are freely  
available at [www.  
openstreetmap.  
org](http://www.openstreetmap.org)

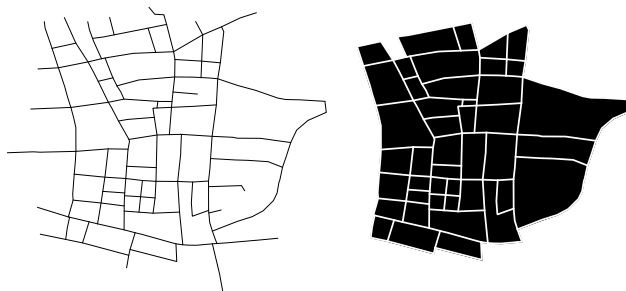


Figure 41: **From the street network to blocks.** Example of a street pattern taken in the neighbourhood of Shibuya in Tokyo (Japan) and the corresponding set of blocks. Note that the block representation does not take into account dead-ends.

different areas, leading to patterns visually different and to cities of different structure. It is thus important to take into account not only the topology of the planar graph — as described by the adjacency matrix — but also the position of the nodes. In order to do that, we propose in this article, a method to characterize this complex object by extracting the ‘fingerprint’ of a street pattern. These fingerprints allow us to define a measure of the distance between two graphs and to construct a classification of cities.

### 13.1 STREETS VERSUS BLOCKS

A major shortcoming of existing classifications is that they are mostly based on the street network. This is however problematic, for two different reasons. First, there is no unambiguous, purely geometrical definition of what a street is: we could define it as the road segment between two intersections, as an almost straight line (up to a certain angular tolerance, see [179]), or we could also follow the actual street names. There is a certain degree of arbitrariness in each of these definitions, and it is not clear how robust a classification based on streets would be. Second, it seems that what is perceived by the human eye of a city map is not coming from streets but from the distribution of the shape, area and disposition of blocks (see Fig. 41).

A natural idea when trying to classify cities is thus to focus on blocks (or cells, or faces) rather than streets. A block can usually be defined without ambiguity as being the smallest area delimited by roads (it has then to be distinguished from a parcel which is a tax related definition). While the information contained in the blocks and the streets are equivalent (up to dead-ends), the information related to the visual aspect of the street network seems to be easier to extract from blocks which are simple geometrical objects — polygons — whose properties are easily measured. The block seems then to be a good candidate for attempting a classification of city patterns.



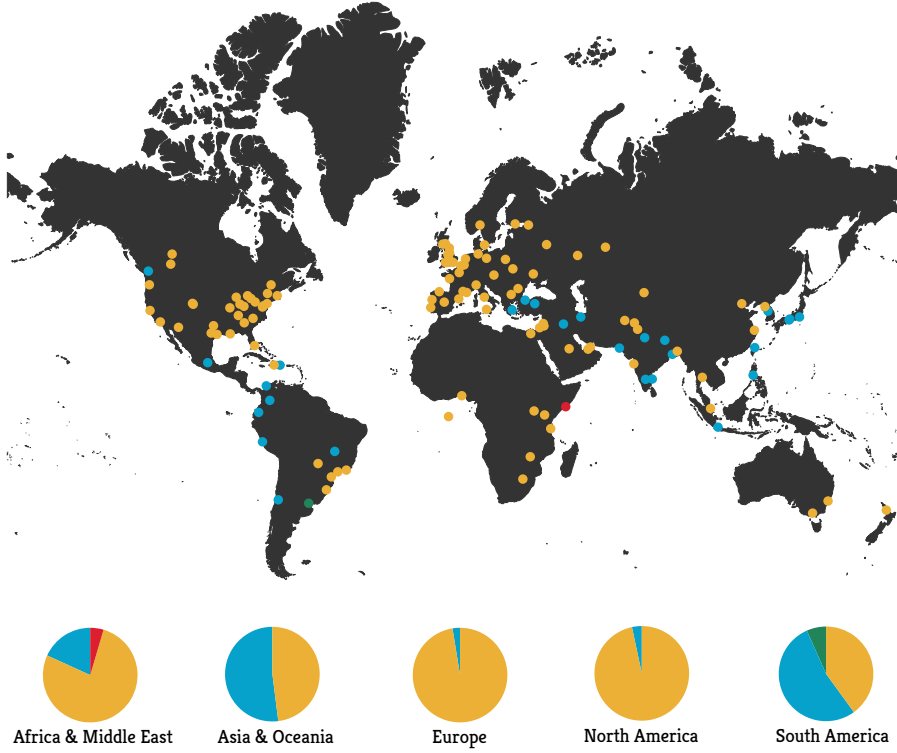


Figure 42: **Location of the cities in our dataset and geographical repartition of the different groups.** The color of the dots indicates in which group the city falls, as defined on Fig 43. On the bottom of the map, the pie charts display the relative importance of the different groups per continent for cities in our dataset (Group 1: 0.8%, Group 2: 20.9%, Group 3: 77.5%, Group 4: 0.8%). We see that the group 3, composed of cities with blocks of various shapes and a slight predominance of larger areas is by far the most represented group in the world.

### 13.2 CHARACTERIZING BLOCKS

Blocks are defined as the cells of the planar graph formed by streets, and it is relatively easy to extract them from a map. We have gathered road networks for 131 major cities across the world, spanning all continents (but Antartica), and their locations are represented on the map Fig. 42. The street networks have been obtained from the OpenStreetMap database, and restricted to the city center using the Global Administrative Areas database (or databases provided by the countries administration). We extracted the blocks from the street network and cleaned undesired features. We end up with a set of blocks, each with a geographical position corresponding to their centroids.

Blocks are polygons and as such can be characterized by simple measures. First, the surface area  $A$  of a block gives a useful indication, and its distribution is an important information about the block pattern. As in [129, 87], we find that for different cities the distributions have

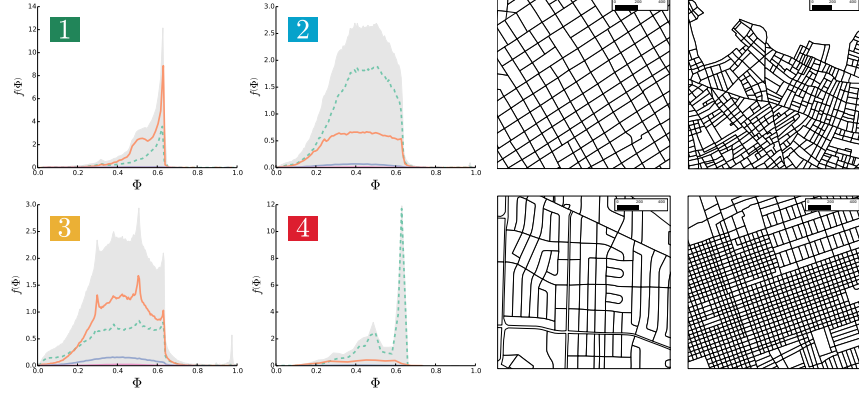


Figure 43: **The four groups.** (Left) Average distribution of the shape factor  $\Phi$  for each group found by the clustering algorithm (Right) Typical street pattern for each group (plotted at the same scale in order to observe differences both in shape and areas). Group 1 (top left): Buenos Aires | Group 2: Athens | Group 3: New Orleans | Group 4: Mogadishu

different shapes for small areas, but display fat tails decreasing as a power law

$$P(A) \sim \frac{1}{A^\tau} \quad (98)$$

with an exponent of order  $\tau \approx 2$  [129, 22, 218, 24]. Although this seemingly universal behaviour gives a useful constraint on any model that attempts at modeling the evolution of cities' road networks, it does not allow to distinguish cities from each other.

A second characterization of a block is through its shape, with the form (or shape) factor  $\Phi$ , defined in the Geography literature in [106] as the ratio between the area of the block and the area of the circumscribed circle  $\mathcal{C}$

$$\Phi = \frac{A}{A_{\mathcal{C}}} \quad (99)$$

The quantity  $\Phi$  is always smaller than one, and the smaller its value, the more anisotropic the block is. There is not a unique correspondence between a particular shape and a value of  $\Phi$ , but this measure gives a good indication about the block's shape in real-world data, where most blocks are relatively simple polygons. The distributions of  $\Phi$  displays important differences from one city to another, and a first naive idea would be to classify cities according to the distribution of block shapes given by  $P(\Phi)$ . The shape itself is however not enough to account for visual similarities and dissimilarities between street patterns. Indeed, we find for example that for cities such as New-York and Tokyo, even if we observe similar distributions  $P(\Phi)$  (see Fig. 44), the visual similarity between both cities's layout is not obvious at all. One reason for this is that blocks can have a similar shape but very different areas: if two

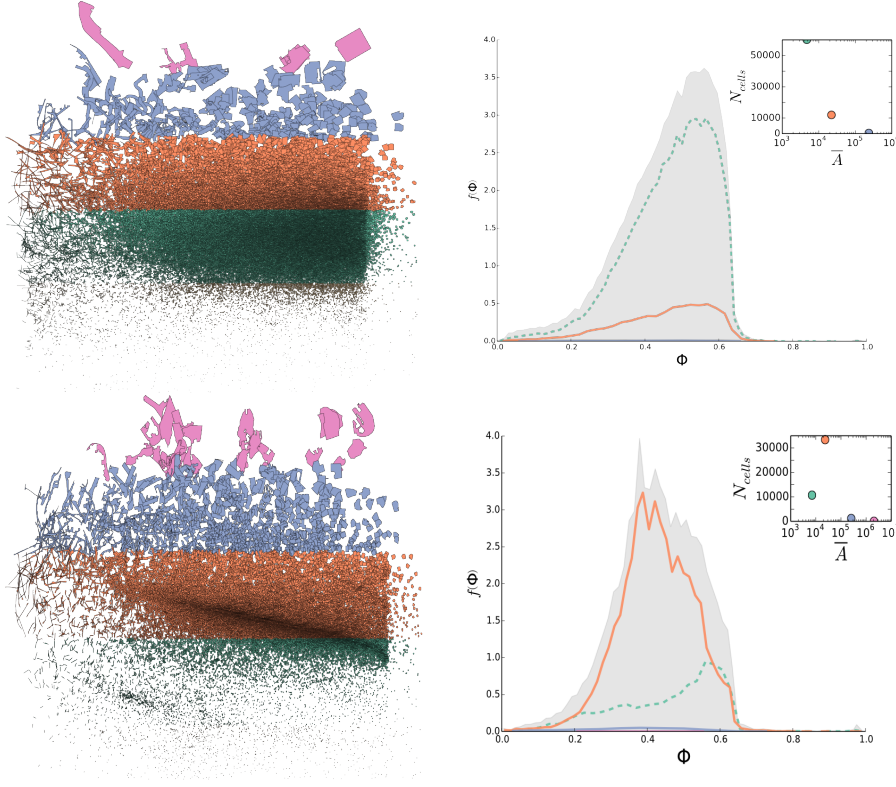


Figure 44: **The fingerprints of Tokyo (top) and New-York, NY (bottom).** (Left) We rearrange the blocks of a city according to their area (y-axis), and their  $\Phi$  value (x-axis). The color of each block corresponds to the area category it falls into. (Right) We quantify this pattern by plotting the distribution of shapes, as measured by  $\Phi$  for each area category, represented by coloured curves. The gray curve is the sum of all the coloured curve and represents the distribution of  $\Phi$  for all cells. As shown in the inset, we see that intermediate area categories dominate the total number of cells, and are thus enough for the clustering procedure.

cities have blocks of the same shape in the same proportion but with totally different areas, they will look different. We thus need to combine the information about both the shape and the area.

In order to construct a simple representation of cities which integrates both area and shape, we rearrange the blocks according to their area (on the y-axis) and display their  $\Phi$  value on the x-axis (Fig. 44). We divide the range of areas in (logarithmic) bins and the color of a block represents the area category to which it belongs. We describe quantitatively this pattern by plotting the conditional probability distribution  $P(\Phi|A)$  of shapes, given an area bin (Fig. 44, right). The colored curves represent the distribution of  $\Phi$  in each area category, and the curve delimited by the gray area is the sum of all these curves and is the

distribution of  $\Phi$  for all cells, which is simply the translation of the well-known formula for probability conditional distribution

$$P(\Phi) = \sum_A P(\Phi|A) P(A) \quad (100)$$

These figures give a ‘fingerprint’ of the city which encodes information about both the shape and the area of the blocks. In order to quantify the distribution of blocks inside a city, and thus the visual aspect of the latter, we will then use  $P(\Phi|A)$  for different area bins. The comparison between these quantities will provide the basis for a classification of street patterns that we propose here.

### 13.3 A TYPOLOGY OF CITIES ACROSS THE WORLD

Two cities will display similar patterns if their blocks have both similar area and shape. In other words, the shape distributions for each area bin should be very close, and this simple idea allows us to propose a distance between street patterns of different cities. More precisely, as one can see on Fig. 44, the number of blocks of area in the range  $[10^3, 10^5]$  (in square meters) dominate the total number of cells, and we will neglect very small blocks (of area  $< 10^3 \text{m}^2$ ) and very large ones (of area  $> 10^5 \text{m}^2$ ). We thus sort the blocks according to their area in two distinct bins

$$\begin{aligned} \alpha_1 &= \left\{ \text{cells} \mid \mathcal{A} \in [10^3, 10^4] \right\} \\ \alpha_2 &= \left\{ \text{cells} \mid \mathcal{A} \in [10^4, 10^5] \right\} \end{aligned}$$

We denote by  $f_\alpha(\Phi)$  the ratio of the number of cells with a form factor  $\Phi$  that lie in the bin  $\alpha$  over the total number of cells for that city. We then define a distance  $d_\alpha$  between two cities  $a$  and  $b$  characterized by their respective  $f_\alpha^a$  and  $f_\alpha^b$

$$d_\alpha(a, b) = \int_0^1 |f_\alpha^a(\Phi) - f_\alpha^b(\Phi)| \, \mathrm{d}\Phi \quad (101)$$

and we construct a global distance  $D$  between two cities by combining all area bins  $\alpha$

$$D(a, b) = \sum_\alpha d_\alpha(a, b)^2 \quad (102)$$

At this point, we have a distance between two cities’ pattern and we measure the distance matrix between all the 131 cities in our dataset,

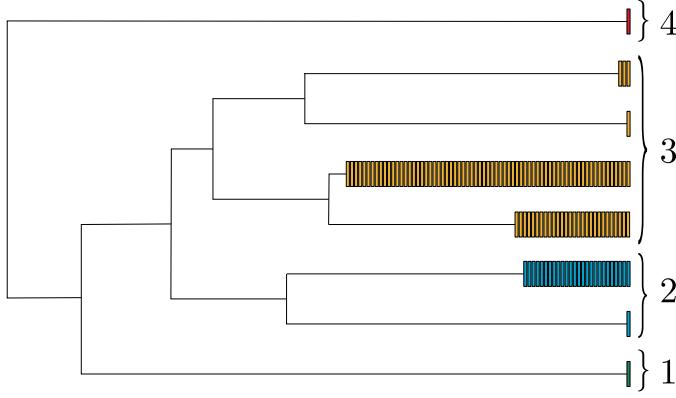


Figure 45: **Dendrogram** We represent the structure of the hierarchical clustering at a given level. Interestingly, 68% of american cities are present in the second largest sub-group of group 3 (fourth from the top). Also, all european cities but Athens are in the largest sub-group of the group 3 (third from top). This result gives a first quantitative grounding to the feeling that European and most American cities are laid out differently.

and perform a classical hierarchical clustering on this matrix [123]. We obtain the dendrogram represented on Fig. 45 and at an intermediate level, we can identify 4 distinct categories of cities, which are easily interpretable in terms of the abundance of blocks with a given shape and with small or large area. On Fig. 43 we show the average distribution of  $\Phi$  for each category and show typical street patterns associated with each of these groups. The main features of each group are the following.

- In the group 1 (comprising Buenos Aires only) we essentially have blocks of medium size (in the bin  $\alpha_2$ ) with shapes that are dominated by the square shape and regular rectangles. Small areas (in bin  $\alpha_1$ ) are almost exclusively squares.
- Athens is a representative element of group 2, which comprises cities with a dominant fraction of small blocks with shapes broadly distributed.
- The group 3 (illustrated here by New Orleans) is similar to the group 2 in terms of the diversity of shapes but is more balanced in terms of areas, with a slight predominance of medium size blocks.
- The group 4 which contains for this dataset the interesting example of Mogadishu (Somalia) displays essentially small, square-shaped blocks, together with a small fraction of small rectangles.

The proportion and location of cities belonging to each group is shown on Fig. 42. Although one should be wary of sampling bias here, it seems that the type of pattern characteristic of the group 3 (various shapes with larger areas) largely dominates among cities in the world. Interestingly, all North American cities (except Vancouver, Canada) are part of

the group 3, as well as all European cities (except Athens, Greece). The composition of the other continents is more balanced between the different groups. Strikingly, we find that at a smaller scale within the group 3 (Fig. 45), all European cities (but Athens) in our sample belong to the same subgroup of the group 3 (the largest one, third from the top on Fig. 45). Similarly, 15 American cities out of the 22 in our dataset belong to the same subgroup of the group 3 (the second largest one, fourth from the top on Fig. 45). Exceptions are Indianapolis (IN), Portland (OR), Pittsburgh (PA), Cincinnati (OH), Baltimore (MD), Washington (DC), and Boston (MA), which are classified with European cities, confirming the impression that these US cities have an european imprint. These results point towards important differences between US and European cities, and could constitute the starting point for the quantitative characterization of these differences [48].

#### 13.4 A LOCAL ANALYSIS

Cities are complex objects, and it is unlikely that an object as simple as the fingerprint can describe all its intricacies. Indeed, cities are usually made of different neighbourhood which often exhibit different street patterns. In Europe, the division is usually clear between the historical center and the more recent suburbs. A striking example of such differences is the Eixample neighbourhood in Barcelona, very distinct from other areas of the city. In order to illustrate this difference, and to show that they also can be captured with our method, we isolate the different Boroughs of New-York, NY: the Bronx, Brooklyn, Manhattan, Queens and Staten Island. We extract the fingerprint of each Borough, as represented on Fig. 46. The fingerprint of New-York (bottom Fig. 44) is indeed the combination of different fingerprints for each of the boroughs. While Staten Island and the Bronx have very similar fingerprints, the others are different. Manhattan exhibits two sharp peaks at  $\Phi \approx 0.3$  and  $\Phi \approx 0.5$  which are the signature of a grid-like pattern with the predominance of two types of rectangles. Brooklyn and the Queens exhibit a sharp peak at different values of  $\Phi$ , also the signature of grid-like patterns with different rectangles for basic shapes.

#### 13.5 DISCUSSION AND PERSPECTIVES

We have introduced a new way of representing cities' road network that can be seen as the equivalent of fingerprints for cities. It seems reasonable to think that the possibility of a classification based on these fingerprints hints at common causes behind the shape of the networks of cities in the same categories. Of course, the present study has limitations: even if the shape of the blocks alone is good enough for the purpose of giving a rough classification of cities, we miss some aspects of the patterns. Indeed, the way the blocks are arranged together lo-

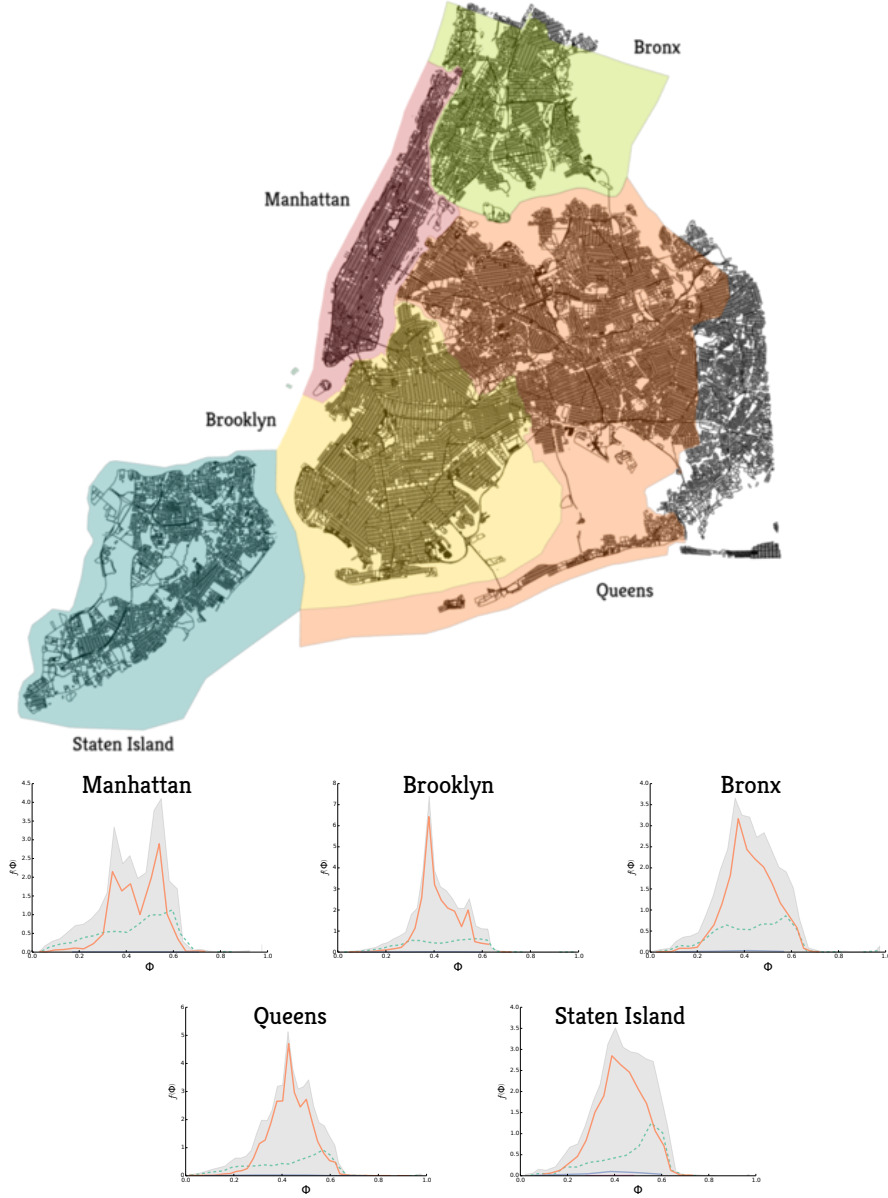


Figure 46: **New-York, NY and its different boroughs** (Top) We represent New York City and its 5 boroughs: the Bronx, Brooklyn, Manhattan, Queens, and Staten Island. (Bottom) The corresponding fingerprints for each borough. Only Staten Island and the Bronx have similar fingerprints and the others are different. In particular, Manhattan exhibits two sharp peaks at  $\Phi \approx 0.3$  and  $\Phi \approx 0.5$  which are the signature of a grid-like pattern with the predominance of two types of rectangles. Brooklyn and the Queens exhibit a sharp peak at different values of  $\Phi$ , signalling the presence of grid-like patterns made of different basic rectangles.



cally should also give some information about the visual aspect of the global pattern. Indeed, many cities are made of neighbourhoods, built at different times, with different street patterns. What is lacking at this point is a systematic, quantitative way to identify and distinguish different neighbourhoods, and to describe their correlation. Indeed, the Boroughs taken as examples in the last section are administrative, arbitrary definitions of a neighbourhood. Reality is however more complex: similar patterns might span several administrative regions, or a given administrative division might host very distinct neighbourhoods. A further step in the classification would thus be to find a method to extract these neighbourhoods, and integrate the spatial correlations between different types of neighbourhoods.

Despite the simplifications that our method entail, we believe that the classification we propose is an encouraging step towards a quantitative and systematic comparison of the street patterns of different cities. This, together with the specific knowledge of architects, urbanists, etc. should lead to a better understanding of the shape of our cities. Further studies are indeed needed in order to relate the various types that we observe to different urban processes. For example, in some cases, small blocks are obtained through a fragmentation process, and their abundance could be related to the age of the city. A large regularity of cell shapes could be related to planning such as in the case of Manhattan for example, but we also know with the example of Paris [24] that a large variety of shapes is also directly related to the effect of a urban modification which does not respect the existing geometry.

Finally, we believe that important empirical progress could be made. A first limitation of the current study is the amount of data that we have. Although 131 cities is a larger number than what is used in most studies, the OpenStreetMap database contains the street layout of many more cities. The more cities we have, the better the classification. We should thus attempt to include more cities.

The second limitation is the use of the administrative definition of cities to delineate the boundaries of the street network. Although it is important to have a large number of cities, it is at least as important to have a set of coherent, similarly defined cities. Administrative definitions, because they are based on political criteria, are completely arbitrary and do not reflect any property of the contained networks. As a result, the chosen boundaries are likely to vary from one country to another, from one city to another. The measures we perform on each of the 131 street patterns are thus, strictly speaking, not comparable. A possible solution would be to use the delineation method proposed by Masucci et al. [153], which is parameter-free and based only on the properties of the street network.



## COST-BENEFIT CONSIDERATIONS IN THE GROWTH OF SPATIAL NETWORKS

---

The following chapter is a reprint of an article, *Emergence of hierarchy in the cost-driven growth of spatial networks*, that was previously published by the author of this thesis with Pablo Jensen and Marc Barthélemy [1].

Our societies rely on various networks for the distribution of energy, information and for transportation of individuals. These networks shape the spatial organization of our societies and their understanding is a key step towards the understanding of the characteristics and the evolution of our cities [28]. Despite their apparent diversity, these networks are all particular examples of a broader class of networks –spatial networks– which are characterised by the embedding of their nodes in space. As a consequence, there is usually a cost associated with a link, leading to particular structures which are now fairly well understood [22], thanks to the recent availability of large sets of data. Nevertheless, the mechanisms underlying the formation and temporal evolution of spatial networks have not been much studied. Different kinds of models aiming at explaining the static characteristics of spatial networks have been suggested previously in quantitative geography, transportation economics, and physics (for a review, see [230]). Concerning the time evolution of spatial networks, a few models only exist to describe in particular the growth of road and rail networks [138, 96, 23, 63], but a general framework is yet to be discovered.

The earliest attempts can be traced back to the economic geography community in the 60s and 70s (A fairly comprehensive review of these studies can be found in [107]). However, due to the lack of available data and computational power, most of the proposed models were based on intuitive, heuristic rules and have not been studied thoroughly. Interestingly, [42] attempts to reproduce railway networks with the same cost-benefits approach that will be adopted in the following.

A more recent trend is that of the optimization models. The common point between all these models is that they try to reproduce the topological features of existing networks, by considering the network as the realisation of the optimum of given quantity (see section IV.E in [22] for an overview). For instance, the hub-and-spoke models [169] reproduce correctly with an optimization procedure the observed hierarchical organization of city pair relations. However, the vast majority of the existing spatial networks do not seem to result from a global optimization, but rather from the progressive addition of nodes and segments resulting

from a local optimization. By modeling (spatial) networks as resulting from a global optimization, one overlooks the usually limited time horizon of planners and the self-organization underlying their formation.

Self-organization of transportation networks has already been studied in transportation engineering [138, 230]. Using an agent-based model including various economical ingredients, the authors of [138] modeled the emergence of the networks properties as a degeneration process. Starting from an initial grid, traffics are computed at each time step and each edge computes its costs and benefits accordingly, using any excess to improve their speed. After several iterations, a hierarchy of roads emerges. Our approach is very different: we start from nodes and we do not specify any initial network. Also, and most importantly, we deliberately do not represent all the causal mechanisms at work in the system. Indeed, the aim of our model is to understand the basic ingredients for emergence of patterns that can be observed in various systems and we thus focus on a single, very general economical mechanism and its consequence on the large-scale properties of the networks.

Concerning spatial networks, as it is the case for many spatial structure, there is a strong path dependency. In other words, the properties of a network at a certain time can be explained by the particular historical path leading to it. It thus seems reasonable to model spatial networks in an iterative way. Some iterative models, following ideas for understanding power laws in the Internet [82] and describing the growth of transportation networks [96] can be found in the literature. In these models, the graphs are constructed via an iterative greedy optimization of geometrical quantities. However, we believe that the topological and geometrical properties of networks are consequences of the underlying processes at stake. At best, geometrical and topological quantities can be a proxy for other –more fundamental– properties: for instance, it will be clear in what follows that the length of an edge can be taken as a proxy for the cost associated with the existence of that edge. Finding those underlying processes is a key step towards a general framework within which the properties of networks can be understood and, hopefully, predicted.

In this respect, cost-benefit analysis (CBA) provides a systematic method to evaluate the economical soundness of a project. It allows one to appreciate whether the costs of a decision will outweigh its benefits and therefore evaluate quantitatively its feasibility and/or suitability. Cost-benefit analysis has only been officially used to assess transport investments since 1960 [61]. However, the concept comes across as so intuitive in our profit-driven economies that it seems reasonable to wonder whether CBA is at the core of the emergent features of our societies such as distribution and transportation systems. If the temporal evolution of spatial networks is rarely studied, arguments mentioning the costs and benefits related to such networks are almost absent from the physics

litterature ([178] is a notable exception, although they do not consider the time evolution of the network.). However, we find it intuitively appealing that in an iterative model, the formation of a new link should—at least locally—correspond to a cost-benefit analysis. We therefore propose here a simple cost-benefit analysis framework for the formation and evolution of spatial networks. Our main goal within this approach is to understand the basic processes behind the self-organization of spatial networks that lead to the emergence of their large scale properties.

#### 14.1 THE MODEL

##### 14.1.1 *Theoretical formulation*

We consider here the simple case where all the nodes are distributed uniformly in the plane (see Methods for detailed description of the algorithm). For a rail network, the nodes would correspond to cities and the network grows by adding edges between cities iteratively; the edges are added sequentially to the graph—as a result of a cost-benefit analysis—until all the nodes are connected. For the sake of simplicity, we limit ourselves to the growth of trees which allows to focus on the emergence of large-scale structures due to the cost-benefit ingredient alone. Furthermore, we consider that all the actors involved in the building process are perfectly rational and therefore that the most profitable edge is built at each step. More precisely, at each time step we build the link connecting a new node  $i$  to a node  $j$  which already belongs to the network, such that the following quantity is maximum

$$R_{ij} = B_{ij} - C_{ij} \quad (103)$$

The quantity  $B_{ij}$  is the *expected benefit* associated with the construction of the edge between node  $i$  and node  $j$  and  $C_{ij}$  is the *expected cost* associated with such a construction. Eq. (103) defines the general framework of our model and we now discuss specific forms of  $R_{ij}$ . In the case of transportation networks, the cost will essentially correspond to some maintenance cost and will typically be proportional to the euclidean distance  $d_{ij}$  between  $i$  and  $j$ . We thus write

$$C_{ij} = \kappa d_{ij} \quad (104)$$

where  $\kappa$  represents the cost of a line per unit of length per unit of time. Benefits are more difficult to assess. For rail networks, a simple yet reasonable assumption is to write the benefits in terms of distance and expected traffic  $T_{ij}$  between cities  $i$  and  $j$

$$B_{ij} = \eta T_{ij} d_{ij} \quad (105)$$

where  $\eta$  represents the benefits per passenger per unit of length. We have to estimate the expected traffic between two cities and for this we will follow the common and simple assumption used in the transportation literature, of having the so-called gravity law [215, 81]

$$T_{ij} = k \frac{M_i M_j}{d_{ij}^a} \quad (106)$$

where  $M_{i(j)}$  is the population of city  $i(j)$ , and  $k$  is the rate associated with the process. We will choose here a value of the exponent  $a > 1$  ( $a < 1$  would correspond to an unrealistic situation where the benefits associated with passenger traffic would increase with the distance). This parameter  $a$  determines the range at which a given city attracts traffic, regardless of the density of cities. The accuracy and relevance of this gravity law is still controversial and improvements have been recently proposed [208, 132]. But it has the advantage of being simple and to capture the essence of the traffic phenomenon: the decrease of the traffic with distance and the increase with population. Within these assumptions, the cost-benefit budget  $R'_{ij} = R_{ij}/\eta$  now reads

$$R'_{ij} = k \frac{M_i M_j}{d_{ij}^{a-1}} - \beta d_{ij} \quad (107)$$

where  $\beta = \frac{\kappa}{\eta}$  represents the relative importance of the cost with regards to the benefits. We will assume that populations are power-law distributed with exponent  $\mu$  (which for cities is approximatively  $\mu \approx 1.1$ , see Methods) and the model thus depends essentially on the two parameters  $a$ , and  $\beta$  (for a detailed description of parameter used in this paper, see the next section). In the following we will be working with fixed values of  $\mu$  and  $a$ . The exact values we choose are however not important as the obtained graphs would have the same qualitative properties.

#### 14.1.2 Simulations

The simulation starts by distributing nodes uniformly in a square. We then attribute to each node a random population distributed according to the power law

$$P_M(x) = \frac{\mu}{x^{\mu+1}} \quad (108)$$

The choice of this distribution is motivated by Zipf's empirical results on city populations [234] (which motivates the choice  $\mu = 1.1$  in our simulations) but also because we can go from a peaked to a broad distribution by tuning the value of  $\mu$ . Indeed, for  $\mu > 2$ , both the first and the second moment of this distribution exist and the distribution

can be considered as peaked. In contrast for  $1 < \mu < 2$ , only the first moment converges and the distribution is broad.

Once the set of nodes is generated, we choose a random node as the root and add nodes recursively until all the nodes belong to the graph. At each time step, the nodes belonging to the graph constitute the set of ‘inactive nodes’, and the other -not yet connected - nodes the ‘active’ nodes. At each time step we connect an active node to an inactive node such that their value of  $R$  defined in Eq. 107 is maximum.

## 14.2 CROSSOVER BETWEEN STAR-GRAPH AND MINIMUM SPANNING TREE

### 14.2.1 Typical scale

The average population is  $\bar{M}$  and the typical inter-city distance is given by  $\ell_1 \sim 1/\sqrt{\rho}$  where  $\rho = N/L^2$  denotes the city density ( $L$  is the typical size of the whole system). The two terms of Eq. 107 are thus of the same order for  $\beta = \beta^*$  defined as

$$\beta^* = k\bar{M}^2\rho^{a/2} \quad (109)$$

In the theoretical discussion that follows, we will take  $k = 1$  for simplicity (but it should not be forgotten in empirical discussions). Another way of interpreting  $\beta^*$  which makes it more practical to estimate from empirical data (see section Discussion), is to say that it is of the order of the average traffic per unit time

$$\beta^* = \langle T \rangle \quad (110)$$

From Eq. 109 we can guess the existence of two different regimes depending on the value of  $\beta$ :

- $\beta \ll \beta^*$  the cost term is negligible compared to the benefits term. Each connected city has its own influence zone depending on its population and the new cities will tend to connect to the most influent city. In the case where  $a \approx 1$ , every city connects to the most populated cities and we obtain a star graph constituted of one single hub connected to all other cities.
- $\beta \gg \beta^*$  the benefits term is negligible compared to the cost term. All new cities will connect sequentially to their closest neighbour. Our algorithm is then equivalent to an implementation of Prim’s algorithm [182], and the resulting graph is a minimum spanning tree (MST).

The intermediate regime  $\beta \simeq \beta^*$  however needs to be elucidated. In particular, we have to study if there is a transition or a crossover between the two extreme network structures, and if we have a crossover

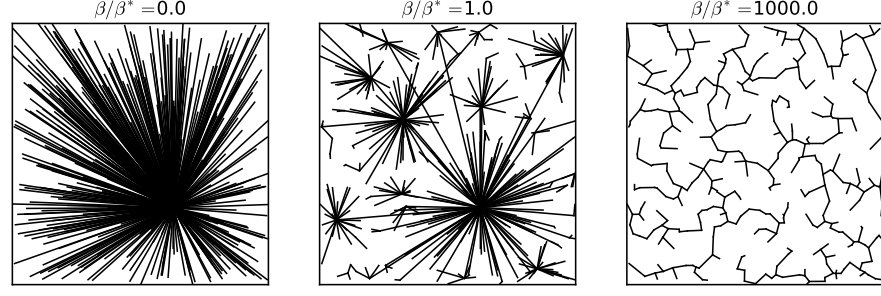


Figure 47: **Simulated graphs.** Graphs obtained with our algorithm for the same set of cities (nodes) for three different values of  $\beta^*$  ( $a = 1.1$ ,  $\mu = 1.1$ , 400 cities). On the left panel, we have a star graph where the most populated node is the hub and on the right panel, we recover the minimum spanning tree.

what is the network structure in the intermediate regime. In the following we answer these questions by simulating the growth of these spatial networks.

#### 14.2.2 Evidence for the crossover

Fig. 47 shows three graphs obtained for the same set of cities for three different values of  $\beta/\beta^*$  ( $a = 1.1$ ,  $\mu = 1.1$ ) confirming our discussion about the two extreme regimes in the previous section. A visual inspection seems to show that for  $\beta \sim \beta^*$  a different type of graph appears, which suggests the existence of a crossover between the star-graph and the MST. This graph is reminiscent of the hub-and-spoke structure that has been used to describe the interactions between city pairs [169, 170]. However, in contrast with the rest of the literature about hub-and-spoke models, we show that this structure is not necessarily the result of a global optimization: indeed, it emerges here as the result of the auto-organization of the system.

The MST is characterised by a peaked degree distribution while the star graph's degree distribution is bimodal, and we therefore choose to monitor the crossover with the Gini coefficient for the degrees defined as in [69]

$$G_k = \frac{1}{2N^2\bar{k}} \sum_{i,j=1}^N |k_i - k_j| \quad (111)$$

where  $\bar{k}$  is the average degree of the network. The Gini coefficient is in  $[0, 1]$  and if all the degrees are equal, it is easy to see that  $G = 0$ . On the other hand, if all nodes but one are of degree 1 (as in the star-graph), a simple calculation shows that  $G = 1/2$ . Fig. 48 displays the evolution of the Gini coefficient versus  $\beta/\beta^*$  (for different values of  $\beta^*$  obtained by changing the value of  $a$ ,  $\mu$  and  $N$ ). This plot shows a smooth variation

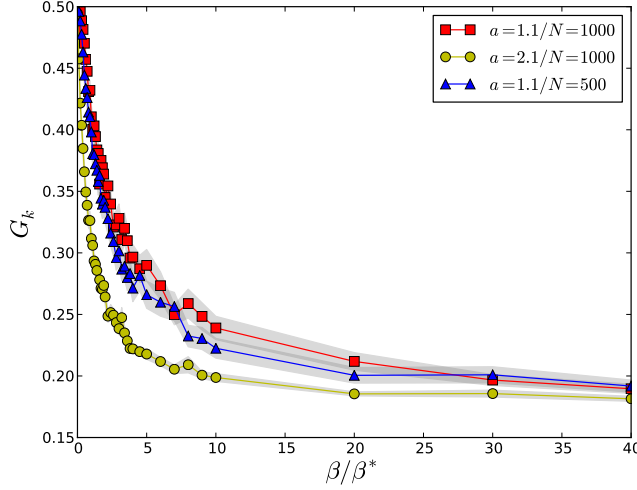


Figure 48: **Gini on node degrees.** Evolution of the Gini coefficient with  $\beta/\beta^*$  for different values of  $\beta^*$ . The shaded area represents the standard deviation of the Gini coefficient. Values decrease from 0.5 in the star-graph regime to below 0.20 in the MST regime.

of the Gini coefficient pointing to a crossover between a star graph and the MST, as one could expect from the plots on Fig. 47 (also, we note that for given values of  $a$ ,  $\mu$  all the plots collapse on the same curve, regardless of the number  $N$  of nodes. However for different values of  $a$  or  $\mu$  we obtain different curves).

Another important difference between the star-graph and the MST lies in how the total length of the graph scales with its number of nodes. Indeed, in the case of the star-graph, all the nodes are connected to the same node and the typical edge length is  $L$ , the typical size of the system the nodes are enclosed in. We thus obtain

$$L_{tot} \sim L N \quad (112)$$

On the other hand, for the MST each node is connected roughly to its nearest neighbour at distance typically given by  $\ell_1 \sim L/\sqrt{N}$ , leading to

$$L_{tot} \sim L \sqrt{N} \quad (113)$$

More generally, we expect a scaling of the form  $L_{tot} \sim N^\tau$  and on Fig. 49 we show the variation of the exponent  $\tau$  versus  $\beta$ . For  $\beta = 0$  we have  $\tau = 1.0$  and we recover the behavior  $L_{tot} \propto N$  typical of a star graph. In the limit  $\beta \gg \beta^*$  we also recover the scaling  $L_{tot} \propto \sqrt{N}$ , typical of a MST. For intermediate values, we observe an exponent which varies continuously in the range  $[0.5, 1.0]$ . This rather surprising behavior is rooted in the heterogeneity of degrees and in the following,

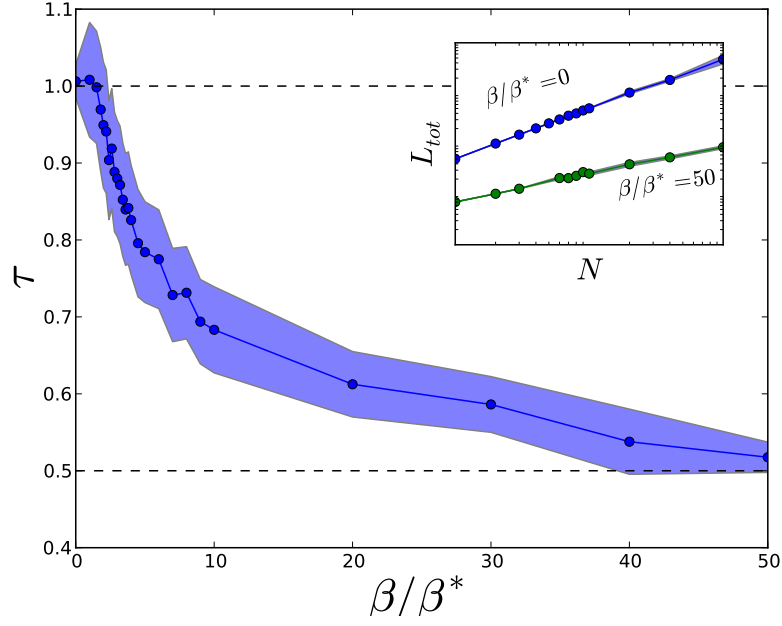


Figure 49: **Star graph to MST transition.** Exponent  $\tau$  versus  $\beta$ . For  $\beta \ll \beta^*$  we recover the star-graph exponent  $\tau = 1$  and for the other extreme  $\beta \gg \beta^*$  we recover the MST exponent  $\tau = 1/2$ . In the intermediate range, we observe a continuously varying exponent suggesting a non-trivial structure. The shaded area represents the standard deviation of  $\tau$ . (Inset) In order to illustrate how we determined the value of  $\tau$ , we represent  $L_{tot}$  versus  $N$  for two different values of  $\beta$ . The power law fit of these curves gives  $\tau$ .

we will show that we can understand this behaviour as resulting from the hierarchical structure of the graphs in the intermediate regime.

It is interesting to note that a scaling with an exponent  $1/2 < \tau < 1$  has been observed [199, 22] for the total number  $\ell_T$  of miles driven by the population (of size  $P$ ) of city scales as  $\ell_T \propto P^\beta$  with  $\beta = 0.66$ . Understanding the origin of those intermediate numbers might thus also give us insights into important features of traffic in urban areas and the structure of cities.

It thus seems that from the point of view of interesting quantities such as the Gini coefficient or the exponent  $\tau$ , there is no sign of a critical value for  $\beta$  and that we are in presence of a crossover and not a transition.

### 14.3 SPATIAL HIERARCHY

The graph corresponding to the intermediate regime  $\beta \approx \beta^*$  depicted on Fig. 47 exhibits a particular structure corresponding to a hierarchical organization, observed in many complex networks [198]. Inspired from the observation of networks in the regime  $\beta/\beta^* \sim 1$ , we define a



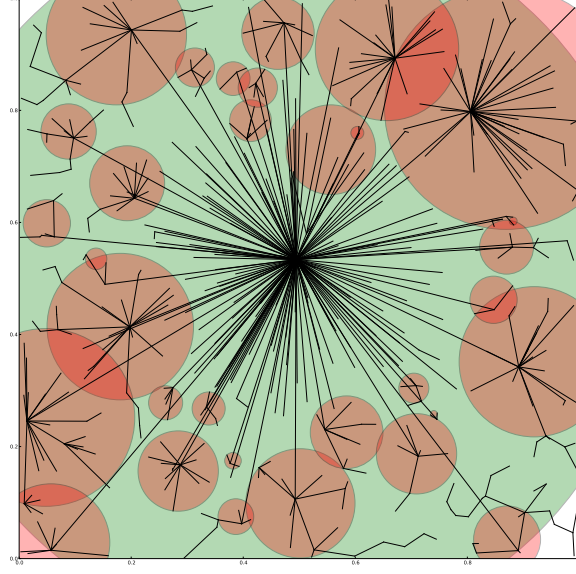


Figure 50: **Influence zones.** Example of a graph where we represent the influence zones for the first two hierarchical levels.

particular type of hierarchy –that we call *spatial hierarchy*– as follows. A network will be said to be spatially hierarchical if:

1. We have a hierarchical network of hubs that connect to nodes less and less far away as one goes down the hierarchy;
2. Hubs belonging to the same hierarchy level have their own influence zone clearly separated from the others'. In addition, the influence zones of a given level are included in the influence zones of the previous level.

The relevance of this new concept of hierarchy in the present context can be qualitatively assessed on Fig. 50 where we represent the influence zones by colored circles, the colors corresponding to different hierarchical levels. In order to go beyond this simple, qualitative description of the structure, we provide in the following a quantitative proof that networks in the regime  $\beta/\beta^*$  exhibit spatial hierarchy.

#### 14.3.1 Distance between hierarchical levels

We propose here a quantitative characterisation of the part (1) in the definition of spatial hierarchy. The first step is to identify the root of the network which allows us to naturally characterize a hierarchical level by its topological distance to the root. We choose the most populated node as the root (which will be the largest hub for  $\beta \ll \beta^*$ ) and we can now measure various quantities as a function of the level in the hierarchy. In Fig. 54, we plot the average euclidean distance  $\bar{d}$  between the different hierarchical levels as a function of the topological distance from the

root node (for the sake of clarity, we also draw next to these plots the corresponding graphs). For reasonably small values of  $\beta/\beta^*$  (i.e. when the graph is not far from being a star-graph), the average distance between levels decreases as we go further away from the root node. This confirms the idea that the graphs for  $\beta/\beta^* \simeq 1$  exhibit a spatial hierarchy where nodes from different levels are getting closer and closer to each other as we go down the hierarchy. Eventually, as  $\beta/\beta^*$  becomes larger than 1, the distance between consecutive levels just fluctuates around  $\ell_1 \sim 1/\sqrt{\rho}$  the average distance between nearest neighbours for a Poisson process, which indicates the absence of hierarchy in the network.

### 14.3.2 Geographical separation of hubs zones

We now discuss the part (2) of the definition of spatial hierarchy, that is to say how the hubs are located in space. Indeed, another property that we can expect from spatially hierarchical graph is that of *geographical separation*.

#### 14.3.2.1 Separation

We say that a graph is geographically separated if the influence zones of every node of a given hierarchical level do not overlap and if they are included in the influence zone of the nodes of the previous level in the hierarchy. Formally, if we designate by  $\mathcal{I}_l^i$  the influence zone of the node  $i$  located at level  $l$  in the hierarchy,  $\mathcal{I}_l = \cup_{i \in l} \mathcal{I}_l^i$  the reunion of all the influence zones for nodes belonging to the level  $n$ . We say that the graph is geographically separated if:

$$\mathcal{I}_l \subset \mathcal{I}_{l+1} \quad \forall l \quad (114)$$

$$\mathcal{I}_l^i \cap \mathcal{I}_l^j = \emptyset \quad \text{if } i \neq j, \forall l \quad (115)$$

The degree of geographical separability of a graph strongly depends on the definition of the influence zone of a node. For instance, if we take the influence zone of a node  $i$  to be the surface of smallest area containing all the nodes connected to  $i$ , it follows that all planar graphs are totally separated. In the context of transportation networks, we expect hubs to radiate up to a certain distance around them, that is to say connect to all the nodes located in a convex shape. We simply define the influence zone of a node  $i$  as the circle centered on the barycenter of  $i$ 's neighbours that belong to the next level, of radius the maximum distance between the barycenter and those points.

Figure 50 is intended to help the reader visualise these influence zones on an example: The green circle represents the influence zone of the root and the red circles the influence zones of the hubs connected to it.

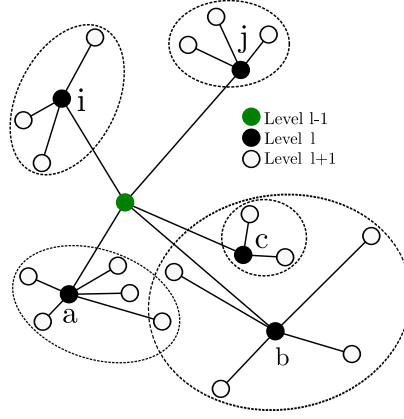


Figure 51: **Influence zones.** Illustration of the influence zones (dotted lines) around several hubs. We have, according to the definition of the separation index,  $S(i, j) = 0$ ,  $0 < S(a, b) < 1$  and  $S(b, c) = 1$ .

One can see that the graph is geographically separated up to a good approximation.

In order to quantify this notion of geographical separability, we define the separation index of the level  $l$  as the average over all the nodes belonging to  $l$  of the separation function. The separation function is equal to 1 if the distance  $d(i, j)$  between the centers of the influence zones of  $i$  and  $j$  is larger than their respective radius (no overlap), and equal to

$$S(i, j) = 1 - \frac{\text{Area of the overlap between } \mathcal{I}_l^i \text{ and } \mathcal{I}_l^j}{\min \left( \text{Area of } \mathcal{I}_l^i, \text{Area of } \mathcal{I}_l^j \right)} \quad (116)$$

One can see that the separation function is equal to 1 if the nodes' influence zones do not overlap at all and 0 if they perfectly overlap (all the influence zones overlapping, like Russian dolls). Therefore, the separation index is equal to 1 if the level  $s$  is perfectly separated and 0 if the influence zones are completely mixed. One can see on Fig. 51 an illustration expliciting the value of the separation index for different situations.

#### 14.3.2.2 Geographical separation in the intermediate regime

We plot the separation index averaged over the all the graph's levels for different values of  $\beta/\beta^*$  on Fig. 52. One can observe on this graph that the separation index reaches values above 0.90 when  $\beta/\beta^* \geq 1$ , which means that the corresponding graphs indeed have a structure with hubs controlling geographically well-separated regions. Obviously, the choice of the shape of the influence zone (which is chosen here to be a disk) strongly impacts the results but the same qualitative behavior will be obtained for any type of convex shapes.

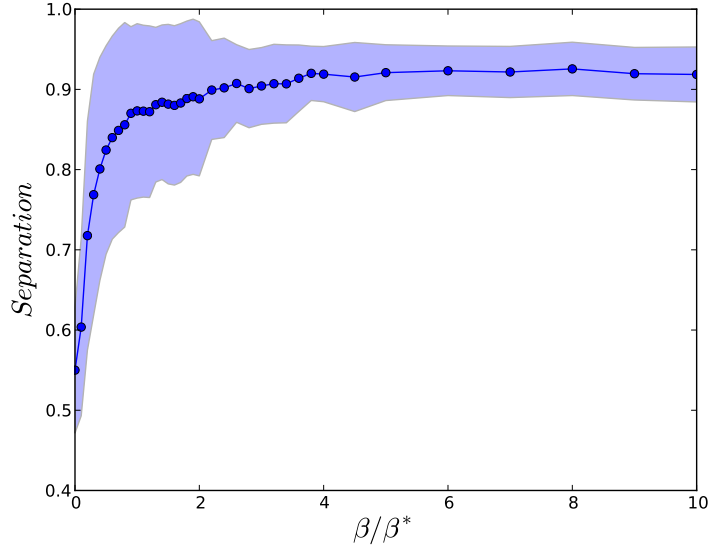


Figure 52: **Separation index.** Separation index averaged over all the graph's level versus  $\beta/\beta^*$ . The shaded area represents the standard deviation.

In conclusion, the graphs produced by our model in the regime  $\beta/\beta^*$  satisfy the two points of the definition. They exhibit a spatially hierarchical structure, characterised by a distance ordering and geographical separation of hubs. We saw earlier that in this regime we have specific, non trivial properties such as  $L_{tot}$  scaling with an exponent depending continuously on  $\beta/\beta^*$ . Using a simple toy model, we will now show that the spatial hierarchy can explain this property.

#### 14.3.3 Understanding the scaling with a hierarchical model

The exponents 1 and 0.5 for the scaling for  $L_{tot}$  with the total number of nodes  $N$  is well-understood. However, it is not clear how we can obtain intermediate values. In the following we show with a simplified model that spatial hierarchy can indeed lead to scaling exponents in the range  $[0.5, 1]$ . We consider the toy model defined by the fractal tree depicted on Fig. 53 for which the distance between the levels  $n$  and  $n + 1$  is given by

$$\ell_n = \ell_0 b^n \quad (117)$$

where  $b \in [0, 1]$  is the scaling factor. Each node at the level  $n$  is connected to  $z$  nodes at the level  $n + 1$  which implies that

$$N_n = z^n \quad (118)$$

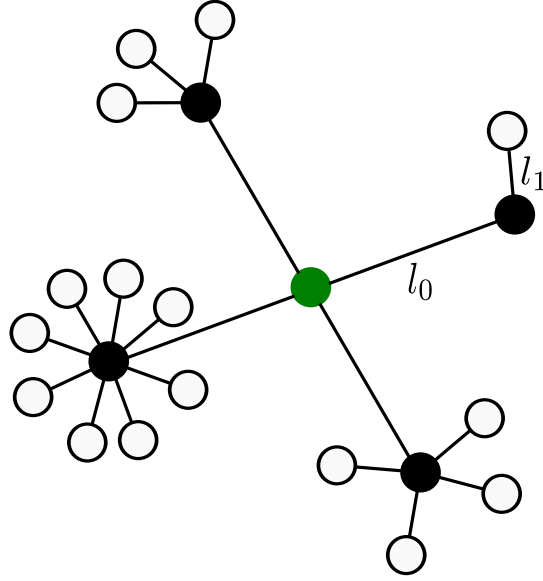


Figure 53: **Fractal toy model.** A schematic representation of the hierarchical fractal network used as a toy model.

where  $z > 0$  is an integer. A simple calculation on this graph shows that in the limit  $z^g \gg 1$ , the total length of the graph with  $g$  levels scales as

$$L_{tot} \sim N^{\frac{\ln(b)}{\ln(z)} + 1} \quad (119)$$

where  $\frac{\ln(b)}{\ln(z)} + 1 \leq 1$  because  $b \leq 1$  and  $z > 1$ . This simple model thus provides a simple mechanism accounting for continuous values of  $\tau$  whose value depends on the scaling factor  $b$ . It provides a simplified picture of the graphs in the intermediate regime  $\beta \simeq \beta^*$  and exhibits the key features of the graphs in this regime: the hub structure reminiscent of the star graph and where the nodes connected to each hub form geographically distinct regions, organized in a hierarchical fashion. It is also interesting to note that the parameter  $z$  can be easily determined from the average degree of the network, and that the parameter  $b$  of the toy model can be related to our model by measuring the decrease of the mean distance between different levels of the hierarchy, as in Fig. 54. By plotting these curves for different values of  $\beta/\beta^*$ , we find that the coefficient of the exponential decays decreases linearly with  $\beta/\beta^*$  and therefore that  $b \sim e^{\beta/\beta^*}$  (However, the comparison only makes sense in the regime  $\beta \sim \beta^*$ , as otherwise the graphs do not exhibit spatial hierarchy).

#### 14.4 EFFICIENCY

Most transportation networks are not obtained by a global optimization but result from the addition of various, successive layers. The question

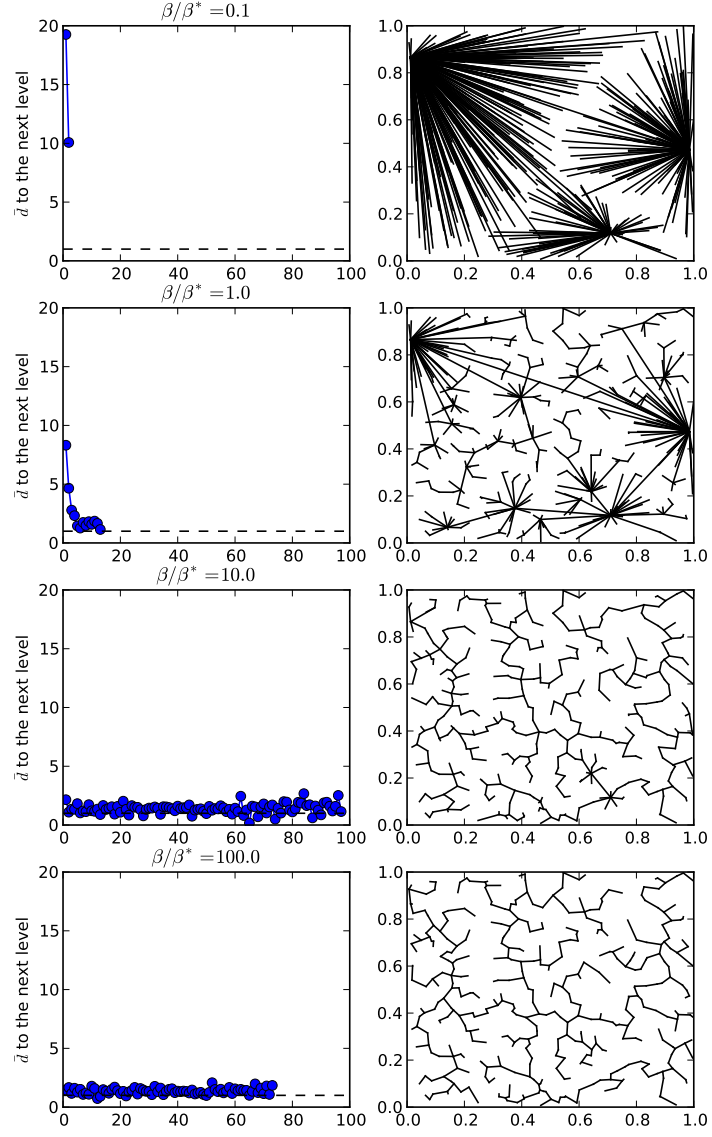


Figure 5.4: **Distance between hierarchy levels.** Left column: Average distance between the successive hierarchy levels for different values of  $\beta/\beta^*$ , next to the corresponding graphs (on the right column). The most populated node is taken as the root node.

of the efficiency of these self-organized systems is therefore not trivial and deserves some investigation. The model considered here allows us to test the effect of various parameters and how efficient a self-organized system can be. In particular, we would like to characterize the efficiency of the system for various values of  $\beta$ . For this, we can assume that the construction cost per unit length is fixed (ie. the factor  $\eta$  in Eq. 104 is constant), and since  $\beta = \frac{\eta}{\kappa}$  a change of value for  $\beta$  is equivalent to a change in the benefits per passenger per unit of length.

A first natural measure of how optimal the network is, is given by its total cost proportional to the total length  $L_{tot}$ : the shorter a network is, the better for the company in terms of building and maintenance costs. In our model, the behaviour of the total cost is simple and expected: for small values of  $\beta/\beta^*$ , the obtained networks correspond to a situation where the users are charged a lot compared to the maintenance cost, and the network is very long ( $L_{tot} \propto N$ ). In the opposite case, when  $\beta/\beta^* \gg 1$  the main concern in building this network is concentrated on construction cost and the network has the smallest total length possible (for a given set of nodes).

The cost is however not enough to determine how efficient the network is from the users' point of view: a very low-cost network might indeed be very inefficient. A simple measure of efficiency is then given by the amount of detour needed to go from one point to another. In other words, a network is efficient if the shortest path on the network for most pairs of nodes is very close to a straight line. The detour index for a pair of nodes  $(i, j)$  is conveniently measured by  $D(i, j)/d(i, j)$  where  $D(i, j)$  is the length of the shortest path between  $i$  and  $j$ , and  $d(i, j)$  is the euclidean distance between  $i$  and  $j$ . In order to have a detailed information about the network, we use the quantity introduced in [11]

$$\phi(d) = \frac{1}{\mathcal{N}(d)} \sum_{\substack{i,j \\ d(i,j)=d}} \frac{D(i,j)}{d(i,j)} \quad (120)$$

where the normalisation  $\mathcal{N}(d)$  is the number of pairs with  $d(i, j) = d$ . We plot this 'detour function' for several values of  $\beta/\beta^*$  on Fig. 55(A). For  $\beta/\beta^* \ll 1$ , the function  $\phi(d)$  takes high values for  $d$  small and low values for large  $d$ , meaning that the corresponding networks are very inefficient for relatively close nodes while being very efficient for distant nodes. On the other hand, for  $\beta/\beta^* \gg 1$  we see that the MST is very efficient for neighboring nodes but less efficient than the star-graph for long distances. Surprisingly, the graphs for  $\beta/\beta^* \sim 1$  exhibit a non trivial behaviour: for small distances, the detour is not as good as for the MST, but not as bad as for the star graph and for long distances it is the opposite. In order to make this statement more precise we compute the average of  $\phi(d)$  over  $d$  (a quantity which has a clear meaning for trees, see [11] for objections to the use of  $\langle \phi(d) \rangle$  as a good efficiency measure in general), and plot it as a function of  $\beta/\beta^*$ . The results

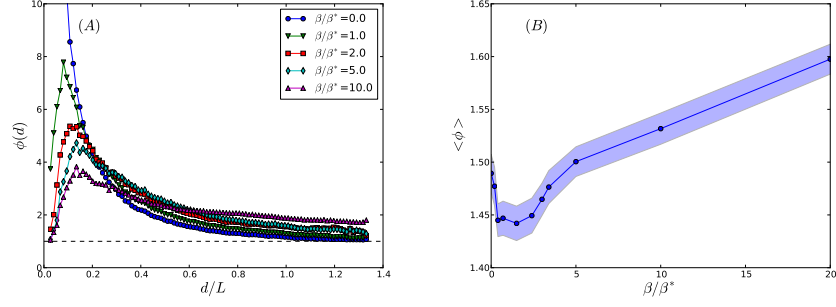


Figure 55: **Detour function.** (Left) Detour function  $\phi(d)$  versus the relative distance between nodes for different values of  $\beta/\beta^*$ . (Right) Average detour index  $\langle \phi \rangle$  for several realisations of the graphs as a function of  $\beta/\beta^*$ . The shaded area represents the standard deviation of  $\langle \phi \rangle$ . This plot shows that there is a minimum for this quantity in the intermediate regime  $\beta \sim \beta^*$ .

are shown in Fig. 55(B) and confirm this surprising behavior in the intermediate regime: we observe a minimum for  $\beta/\beta^* \sim 1$ . In other words, there exists a non trivial value of  $\beta$ , i.e. a value of the benefits per passenger per unit of length, for which the network is optimal from the point of view of the users.

The existence of such an optimum is far from obvious and in order to gain more understanding about this phenomenon, we plot the Gini coefficient  $G_l$  relative to the length of the edges between nodes in Fig. 56. We observe that the Gini coefficient peaks around  $\beta/\beta^* = 1$ , which means that in this regime, the diversity in terms of edge length is the highest. The large diversity of lengths explains why the network is the most efficient in this regime: indeed long links are needed to cover large distances, while smaller links are needed to reach efficiently all the nodes. It is interesting to note that this argument is similar to the one proposed by Kleinberg [126] in order to explain the existence of an optimal delivery time in small-world networks.

#### 14.5 DISCUSSION

We have presented a model of a growing spatial network based on a cost-benefit analysis. This model allows us to discuss the effect of a local optimization on the large-scale properties of these networks. First, we showed that the graphs exhibit a crossover between the star-graph and the minimum spanning tree when the relative importance of the cost increases. This crossover is characterized by a continuously varying exponent which could give some hints about other quantities observed in cities such as the total length travelled by the population. Secondly, we showed that the model predicts the emergence of a spatial hierarchical structure in the intermediate regime where costs and benefits are of the same order of magnitude. We showed that this spatial hierarchy can



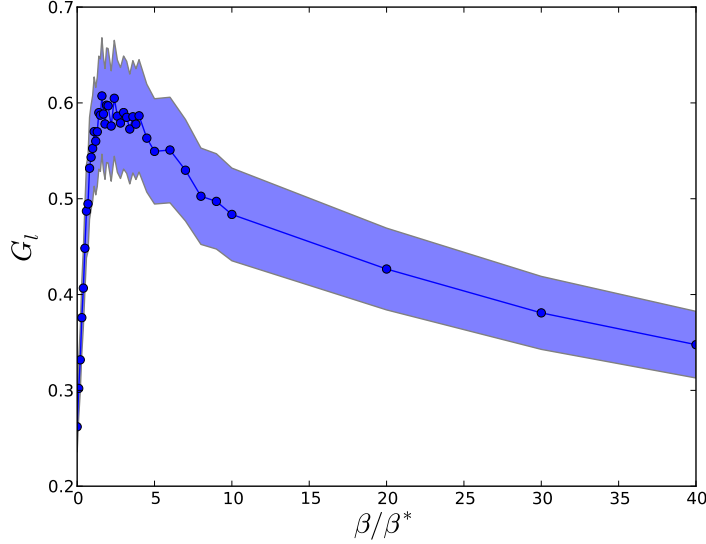


Figure 56: **Gini on the length.** Evolution of the Gini coefficient for the length versus  $\beta/\beta^*$  (for different values of  $\beta^*$ ). The shaded area represents the standard deviation.

explain the non trivial behaviour of the total length versus the number of nodes. Finally, this model shows that in the intermediate regime the vast diversity of links lengths entails a large efficiency, an aspect which could of primary importance for practical applications.

An interesting playground for this model is given by railways and we can estimate the value of  $\beta/\beta^*$  for these systems. In some cases, we were able to extract the data from various sources (in particular financial reports of railway companies) and the results are shown in Table 1. We estimate for different real-world networks, including some of the oldest railway systems,  $\beta$  using its definition (total maintenance costs per year divided by the total length and by the average ticket price per km). In order to estimate  $\beta^*$  we use Eq.110 in the following way

$$\beta^* \simeq \frac{T_{tot}}{L_{tot}} \quad (121)$$

where  $T_{tot}$  is the total travelled length (in passengers·kms/year) and  $L_{tot}$  is the total length of the network under consideration. Remarkably, the computed values for the ratio  $\beta/\beta^*$  shown in Table 10. are all of the order of 1 (ranging from 0.20 to 1.56). In the framework of this model, this result shows that all these systems are in the regime where the networks possess the property of spatial hierarchy, suggesting it is a crucial feature for real-world networks. We note that in our model, the value of  $\beta/\beta^*$  is given exogeneously, and it would be extremely interesting to understand how we could construct a model leading to this value in an endogeneous way.

Country	$T_{tot}$ (kms/year)	$L_{tot}$ (kms)	Maintenance (euros/year)	Ticket price euros	$\beta/\beta^*$
France	$88.1 \cdot 10^9$	29,901	$2.10 \cdot 10^9$	0.12	<b>0.20</b>
Germany	$79.2 \cdot 10^9$	37,679	$7.50 \cdot 10^9$	0.30	<b>0.32</b>
India	$978.5 \cdot 10^9$	65,000	$3.00 \cdot 10^9$	0.01	<b>0.31</b>
Italy	$40.6 \cdot 10^9$	24,179	$4.30 \cdot 10^9$	0.20	<b>0.53</b>
Spain	$22.7 \cdot 10^9$	15,064	$3.16 \cdot 10^9$	0.11	<b>1.26</b>
Switzerland	$18.0 \cdot 10^9$	5,063	$2.03 \cdot 10^9$	0.17	<b>0.66</b>
United Kingdom	$62.7 \cdot 10^9$	16,321	$12 \cdot 10^9$	0.16	<b>1.19</b>
United States	$17.2 \cdot 10^9$	226,427	$2.96 \cdot 10^9$	0.11	<b>1.56</b>

Table 10: **Empirical estimates for  $\beta/\beta^*$ .** Table giving the total ride distance (in km), the total network length (in km), the total annual maintenance expenditure (in euros per year) and the average ticket price (in euros per km). All the given values correspond to the year 2011. From these data we compute the experimental values of  $\beta$ ,  $\beta^*$  and their ratio (data obtained from various sources such as financial reports of railway companies)

There are also several directions that seem interesting. First, various forms of cost and benefits functions could be investigated in order to model specific networks. In particular, there are several choices that can be taken for the expected traffic. In this paper we limited ourselves to estimate the traffic as a direct traffic from a node  $i$  to a node  $j$ , but it is likely that part of the traffic will come from other nodes. In order to take this into account, we think that the following extensions are probably interesting:

1. A given city (denoted by 0 with population  $M_0$ ) plays a particular role in the network (the capital city in a relatively small country, for example). In that case it is beneficial to be close to that city through the network and we write

$$R_{ij}^{(1)} = (1 - \lambda) \frac{M_i M_j}{d_{ij}^{a-1}} + \lambda \frac{M_i M_0}{(D_{0j} + d_{ij})^{a-1}} - \beta d_{ij} \quad (122)$$

where  $\lambda \in [0, 1]$  is a coefficient weighing the relative importance of the traffic coming from the particular city.

2. The most general case where all the network-induced traffic are taken into account. We then consider

$$R_{ij}^{(2)} = \sum_{k \neq i} \frac{M_i M_k}{(D_{kj} + d_{ij})^{a-1}} - \beta d_{ij} \quad (123)$$

Other ingredients such as the presence of different rail companies, or the difference between a state-planned network and a network built by private actors, etc, could easily be implemented and the corresponding models could possibly lead to interesting results.

More importantly, we limited ourselves here to trees in order to focus on the large-scale consequences of the cost-benefit mechanism. Further studies are needed in order to uncover the mechanisms of formation of loops in growing spatial networks and we believe that the model presented here might represent a suitable modeling framework.

Finally, it seems plausible that the general cost-benefit framework introduced at the beginning of the article could be applied to the modelling of systems besides transportation networks. We believe it captures the fundamental features of spatial network while being versatile enough to model the growth of a great diversity of systems shaped by space.



The following chapter is a reprint of an article, *Scaling in transportation networks*, that was previously published by the author of this thesis with Camille Roth and Marc Barthélemy [4].

Almost 200 subway systems run through the largest agglomerations in the world and offer an efficient alternative to congested road networks in urban areas. Previous studies have explored the topological and geometrical static properties of these transit systems [67, 137], as well as their evolution in time [195]. However, subways are not mere geometrical structures growing in empty space: they are usually embedded in large, highly congested urban areas and it seems plausible that some properties of these systems find their origin in the interaction with the city they are in. Previous studies [136, 230] have shown that the growth and properties of transportation networks are tightly linked to the characteristics of urban environment. Levinson [136] for instance, showed that rail development in London followed a logic of both ‘induced supply’ and ‘induced demand’. In other words, while the development of rail systems within cities answers a need for transportation between different areas, this development also has an impact on the organisation of the city. Therefore, while the growth of transportation cannot be understood without considering the underlying city, the development of the city cannot be understood without considering the transportation networks that run through it. As a result, the subway system and the city can be thought as two systems exhibiting a symbiotic behaviour. Understanding this behaviour is crucial if we want to get a deeper understanding of how the city grows and how the mobility patterns organise themselves in urban environments.

At a different scale, railway networks answer a need for fast transportation between different urban centers. We therefore expect their properties to be linked to the characteristics of the underlying country. The model of growth presented in Chapter 14 relates the existence of a given line to the economical and geographical features of the environment. An interesting question is thus to know whether subways and railway networks behave in the same way, but at different scales. In other words, we are interested to know whether subways are merely scaled down railway networks, or whether they are fundamentally different objects, following different growth mechanisms.

In the spirit of the model proposed in the previous Chapter, we propose here a large-scale framework which relates structural and economical properties of subway and railway networks. Although many studies [121, 67, 137] explore the interplay between regional characteristics

and the structure of transportation networks, a simple picture relating the network's most basic quantities and the region's properties is still lacking. It has been found that several biological and man-made systems exhibit allometric scaling relationships between the output of processes and size. These relationships are hints that very general processes are at stake in the growth of these systems, and a first step towards their understanding is to uncover these processes [20, 3]. In the spirit of what has recently been done for cities [3], we try in the following to understand the way subways and railway networks scale with some of the substrates' most basic attributes: population, surface area and wealth.

We believe this should lay the foundations for more specific and involved discussions.

As a result, we are able to relate the total ridership, the number of stations, the length of the network to socio-economical features of the environment. We find that these relations are in good agreement with the data gathered for 138 subway systems and 58 railway networks across the world. In particular, we show that even if the main mechanisms are the same, the difference of scale at which both systems operate is responsible for their different behavior.

*Data for 138  
subways accross  
the world were  
collected on  
Wikipedia, and  
cross-referenced  
with the  
operators' data  
when possible*

#### 15.1 FRAMEWORK

A transportation network is at least characterized by its total number of nodes (which are here train or subway stations), its total length, and the total (yearly) ridership. On the other hand, a city (or a country in the railway case) is characterized by its area, its population and its GDP. Because transportation systems do not grow in empty space, but result from multiple interactions with the substrate, an important question is how network characteristics and socio-economical indicators relate to each other. Naturally, cost-benefit analysis seems to be the appropriate theoretical framework. While this approach has already been developed in the context of the growth of railway networks [42, 1], these studies considered an iterative growth: at each step an edge  $e$  is built such that the cost function

$$Z_e = B_e - C_e \quad (124)$$

is maximum. The quantity  $B_e$  is the expected benefit and  $C_e$  the expected cost of  $e$ . In the following, we consider networks after they have been built, and we assume that they are in a 'steady-state' for which we can write a cost function of the form

$$Z = \sum_e Z_e = B - C \quad (125)$$

where  $B$  is the total expected benefits and  $C$  the total expected costs, now operating costs (mainly maintenance costs). We further assume

that, during this steady-state, operating costs are balanced by benefits. In other words

$$Z \approx 0 \quad (126)$$

Indeed, because lines and stations cost money to be maintained, we expect the network to adapt to the way it is being used. Therefore we can reasonably expect that at first order the cost of operating the system is compensated by the benefits gained from its use. In the following we will apply this general framework to subway and railway networks in order to determine the behavior of various quantities with respect to population and GDP.

## 15.2 SUBWAYS

In the case of subways, the total benefits in the steady-state are simply connected to the total ridership  $R$  and the ticket price  $f$  over a given period of time. The costs, on the other hand, are due to the maintenance costs of the lines and stations, so that we can write (for a given period of time)

$$Z_{sub} = Rf - \epsilon_L L - \epsilon_S N_s \quad (127)$$

where  $L$  is the total length of the network,  $\epsilon_L$  the maintenance cost of a line per unit of length,  $N_s$  the total number of stations and  $\epsilon_S$  the maintenance cost of a station (for a given period time).

It is usually difficult to estimate the ridership of a system given its characteristics and those of the underlying city. Due to the importance of such estimates for planning purposes, the problem of estimating the number of boardings per station given the properties of the area surrounding the stations has been the subject of numerous studies [154, 127]. Here we are interested in the dependence of global, average behaviours of the ridership on the network and the underlying city. Very generally, we write that the number  $R_i$  of people using the station  $i$  will be a function of the area  $C_i$  serviced by this station — the ‘coverage’ [67] — and of the population density  $\rho = \frac{P}{A}$  in the city

$$R_i = \xi_i C_i \rho \quad (128)$$

where  $\xi_i$  is a random number of order one representing the ratio of people covered who use the subway. The main difficulty is in finding the expression of the coverage. It depends, a priori, on local particularities such as the accessibility of the station, and should thus vary from one station to another. We take here a simple approach and assume that on average

$$C_i \sim \pi d_0^2 \quad (129)$$

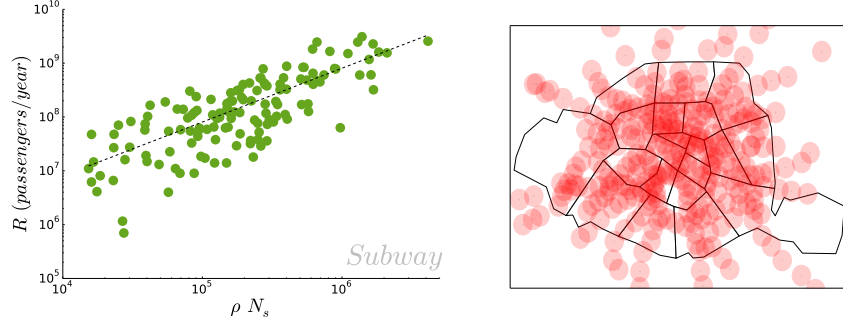


Figure 57: **(Subway) The relationship between ridership and coverage**  
 (Left) We plot the total yearly ridership  $R$  as a function of  $\rho N_s$ . A linear fit on the 138 data points gives  $R \approx 800 \rho N_s$  ( $R^2 = 0.76$ ) which leads to a typical effective length of attraction  $d_0 \approx 500$  m per station. (Right) Map of Paris, France with each subway station represented by a red circle of radius 500 m.

where  $d_0$  is the typical size of the attraction basin of a given station. If we assume that it is constant, the total ridership can be written as

$$R = \sum_i R_i \sim \bar{\xi} \pi d_0^2 \rho N_s \quad (130)$$

where  $\bar{\xi} = \frac{1}{N_s} \sum_i \xi_i$  is of the order of 1.

We gathered the relevant data for 138 metro systems across the world, which we cross-verified when possible with the data given by network operators. While the number of stations, the number of lines, total length of the networks and ridership are relatively straightforward to define, the choice of population and city area is more subtle. Indeed, most subway systems span an area greater than the city core, and the relevant area therefore lies somewhere between the city core's area and the total urbanized area. We chose to use the population and surface area data for urbanized areas provided by Demographia.

We plot the ridership  $R$  as function of  $N_s \rho$  on Fig. 57 and observe that the data is consistent with a linear behavior. We measure a slope of  $800 \text{ km}^2/\text{year}$  which gives an estimate for  $d_0$

$$d_0 \approx 500 \text{ m} \quad (131)$$

We illustrate this result on Fig. 57 by representing the subway stations of Paris each with a circle of radius 500 m.

So far, the distance  $d_0$  appears here an intrinsic feature of user's behaviors: it is the maximal distance that an individual would walk to go to a subway station.

The average interstation distance  $\ell_1$  is another distance characteristic of the subway system. Rigorously, this distance depends on the average



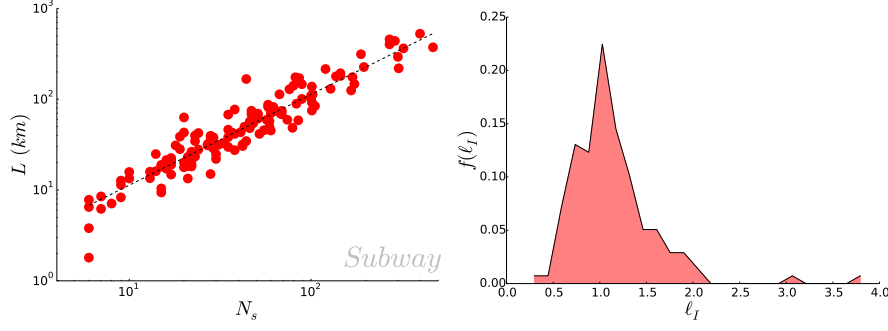


Figure 58: **(Subway) Relation between the length and the number of stations** (Left) Length of 138 subway networks in the world as a function of the number of stations. A linear fit gives  $L \sim 1.13 N_s$  ( $R^2 = 0.93$ ) (Right) Empirical distribution of the inter-station length. The average interstation distance is found to be  $\bar{\ell}_1 \approx 1.2\text{km}$  and the relative standard deviation is approximately 440 m

degree  $\langle k \rangle$  of the network so that  $\ell_1 = \frac{2L}{N_s \langle k \rangle}$ . It has however been found that for the 13 largest subway systems in the world,  $\langle k \rangle \in [2.1, 2.4]$ , so that we can reasonably take  $\langle k \rangle / 2 \approx 1$  and thus

$$\ell_1 \simeq \frac{L}{N_s} \quad (132)$$

The interstation distance depends in general on many technological and economical parameters, but we expect that for a properly designed system it will match human constraints. Indeed, if  $d_0 \ll \ell_1$ , the network is not dense enough and in the opposite case  $d_0 \gg \ell_1$ , the system is not economically interesting. We can thus reasonably expect that the interstation distance fluctuates slightly around an average value given by twice the typical station attraction distance  $d_0$

$$d_0 = \frac{\ell_1}{2} = \frac{L}{2N_s} \quad (133)$$

It follows from this assumption that the interstation distance is constant and independent from the population size. We plot on Fig. 58 the total length of subway networks as a function of the number of stations. The data agrees well with a linear fit  $L \sim 1.13 N_s$  ( $r^2 = 0.93$ ). We also plot on Fig. 58 the histogram of the inter-station length, showing that the interstation distance is indeed narrowly distributed around an average value  $\bar{\ell}_1 \approx 1.2\text{km}$  with a variance  $\sigma \approx 400\text{m}$ , consistently with the value found above for  $d_0 \approx 500\text{m}$ . The outliers are San Francisco, whose subway system is more of a suburban rail service and Dalian, a very large city whose metro system is very young and still under development.

As a result of the previous argument, we can express  $\ell_1$  in terms of the systems characteristics. Indeed, the total ridership now reads

$$R \sim \bar{\xi} \pi \rho \frac{L^2}{N_s} \quad (134)$$

If we assume to be in the steady-state  $Z_{sub} \approx 0$ , using the results from Eqs. (127,134), we find that the total length of the network and the number of stations are linked at first order in  $\epsilon_s/\epsilon_L$  by

$$L \sim \left( \frac{4\epsilon_L}{\pi \bar{\xi} f \rho} + \frac{\epsilon_s}{\epsilon_L} \right) N_s \quad (135)$$

and that the interstation distance reads

$$\ell_1 = \frac{4\epsilon_L}{\pi \bar{\xi} f \rho} + \frac{\epsilon_s}{\epsilon_L} \quad (136)$$

This relation implies that the interstation distance increases with an increased station maintenance cost, and decreases with increased line maintenance costs, density and fare. We thus see that the adjustment of  $\ell_1$  to match  $2d_0$  can be made through the fare price (or subsidies by the local authorities or national government). At this point, it would be interesting to get reliable data about the maintenance costs and fare for subway systems in order to pursue in this direction and test the accuracy of this prediction.

So far, we have a relation between the total length and the number of stations, but we need another equation in order to compute their value. Intuitively, it is clear that the number of stations — or equivalently the total length — of a subway system is an increasing function of the wealth of the city. We assume a simple, linear relation of the form

$$N_s = \beta \frac{G}{\epsilon_s} \quad (137)$$

where  $G$  is the city's Gross Metropolitan Product, and  $\beta$  the fraction of the city's wealth invested in public transportation. On Fig. 59 (left) we plot the number of stations of different metro systems around the world as a function of the Gross Metropolitan Product of the city. A linear fit agrees relatively well with the data ( $R^2 = 0.73$ , dashed line), and gives  $\frac{\epsilon_s}{\beta} \approx 10^{10}$  dollars/station. However, the dispersion around the linear average behaviour is important: more specific data is needed in order to investigate whether differences in the construction costs and investments (or the age of the system) can, alone, explain the dispersion.

Finally, we now consider the number of different lines with distinct tracks. A natural question is how the number of lines  $N_{lines}$  scales with

*The cities' GDP  
per capita was  
retrieved for 114  
cities from  
Brooking's Global  
MetroMonitor.*

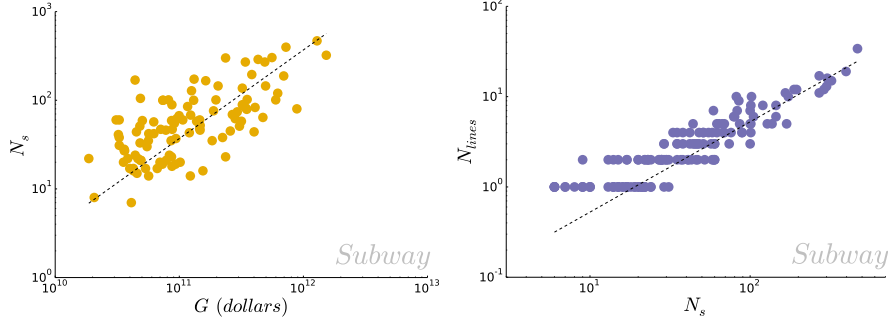


Figure 59: **(Subway) Size of the subway system and city's wealth** We plot the number of stations for the different subway systems in the dataset as a function of the Gross Metropolitan Product of the corresponding cities (obtained for 106 subway systems). A linear fit (dashed line) gives  $N_s = 2.51 \cdot 10^{-10} G$  ( $R^2 = 0.73$ ). **(Subway) Number of lines and number of stations** We plot the number of metro lines  $N_{lines}$  as a function of the number of stations  $N_s$ . A linear fit on the 138 data points gives  $N_{lines} \approx 0.053 N_s$  ( $R^2 = 0.94$ ), or, in other words, metro lines contain on average 19 stations.

the number stations  $N_s$ , that is to say whether lines get proportionally smaller, larger or the same with the size of the whole system. We plot the number of lines as a number of stations on Fig. 59 and find that the data agree with a linear relationship between both quantities ( $R^2 = 0.93$ , see the dashed black line). In other words, the number of stations per line is distributed around a typical value of 19, whatever the size of the system.

### 15.3 RAILWAY NETWORKS

We start by discussing an important difference between railway and subway networks. In the subway case, the interstation distance is such that it matches human constraints:  $\ell_1 \sim 2d_0$  where  $d_0$  is the typical distance that one would walk to reach a subway station. For the railway network, the logic is however different: while subways are built to allow people to move within a dense urban environment, the purpose of building a railway is to connect different cities in a country. In addition, due to the long distance and hence high costs, it seems reasonable to assume that each station is connected to its closest neighbour. In this respect, the railway network appears as a planar graph connecting randomly distributed nodes in the plane in an economical way. If we assume that a country has an area  $A$  and  $N_s$  train stations, the typical distance between nearest stations will be

$$\ell_N = \sqrt{\frac{A}{N_s}} \quad (138)$$

*Data about ridership, network length were easily retrievable for more than 100 countries from the UIC Railis 2011 database.*

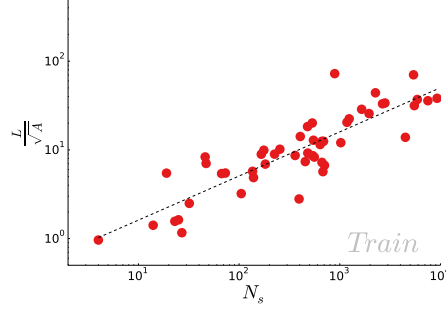


Figure 60: **(Train) Total length and number of stations** Total length of the railway network  $L$  rescaled by the typical size of the country  $\sqrt{A}$  as a function of the number of stations  $N_s$ . The dashed line shows the best power-law fit on the 50 data points with an exponent  $0.50 \pm 0.08$  ( $R^2 = 0.87$ ).

The total length  $L \sim N_s \ell_N$  is then given by

$$L \sim \sqrt{A} N_s \quad (139)$$

In order to test this relation for different countries, we plot the adimensional quantity  $\frac{L}{\sqrt{A}}$  as a function of the number of stations  $N_s$  on Fig. 60. A power law fit gives an exponent  $0.50 \pm 0.08$  ( $R^2 = 0.87$ ), which is consistent with the previous argument.

At this point, we have a relation between  $L$  and  $N_s$ , but we need to find the expressions for the other quantities. There are other differences with the subway system. First, due to the distances involved, the ticket price usually depends on the distance travelled and we will denote by  $f_L$  the ticket price per unit distance. The relevant quantity for benefits is therefore not the raw number of passengers—as in subways—, but rather the total distance travelled on the network  $T$ . Also, again due to the long distances spanned by the network, the costs of stations can be neglected as a first approximation, and we get for the budget the following expression

$$Z_{train} \simeq T f_L - \epsilon_L L \quad (140)$$

In the steady-state regime  $Z_{train} \approx 0$  — or in other words, the revenue generated by the network use must be of the order of the total maintenance costs [1] (see Chapter 14 — we find that

$$T \sim \frac{\epsilon_L}{f_L} L \quad (141)$$

The number of stations was more difficult to find. We had to use various data sources, mainly scrapping the operators' ticket booking websites.

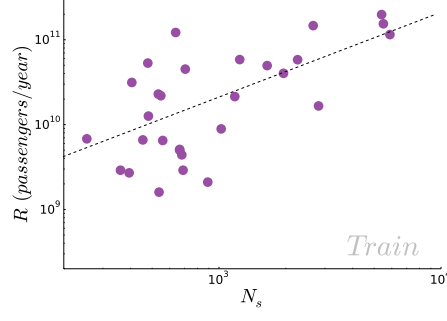


Figure 61: **(Train) Ridership and number of stations** The total yearly ridership  $R$  of the railway networks as a function of the number of stations. A linear fit on the 47 data points gives  $R \sim 7.0 \cdot 10^8 N_s$  ( $R^2 = 0.86$ )

In addition, if we assume that the order of magnitude of a trip is given by  $\ell_N$ , the total travelled length is simply proportional to the ridership  $T \sim \ell_N R$  leading to

$$R \sim \frac{\epsilon_L N_s}{f_L} \quad (142)$$

We thus plot the total daily ridership  $R$  as a function of the total number of stations  $N_s$  (figure 61), and despite the small number of available data points, a linear relationship between these both quantities seems to agree with empirical data on average ( $R^2 = 0.86$ ). This result should be taken with caution, however, due to the important dispersion that is observed around the average behaviour, and the small number of observations.

According to the previous result, the total length and the number of stations are related to each other. We now would like to understand what property of the underlying country determines the total length of the network. That is to say, why networks are longer in some countries than in others. As in subway systems, economical reasons seem appealing. Indeed, the railway networks of some large african countries such as Nigeria are way smaller than that of countries such as France or the UK of similar surface areas. A priori, when estimating the cost of a railway network, one should take into account both the costs of building lines and the stations. However, as stated above, considering the distances involved, the cost of building a station is negligible compared to that of building the actual lines. We thus can reasonably expect to have

$$L \sim \frac{\alpha G}{\epsilon_L} \quad (143)$$

where  $G$  is here the country's Gross Domestic Product (GDP) used as an indicator of the country's wealth, and  $\alpha < 1$  the ratio of the GDP invested in railway transportation. We plot  $L$  as a function of  $G$

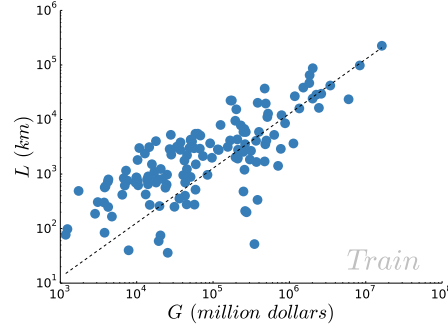


Figure 62: **(Train) Total length of the network and wealth** Total length of the railway network  $L$  as a function of the country GDP  $G$ . The dashed line shows the best linear fit on the 138 data points which gives  $\epsilon_L/\alpha \approx 10^4$  dollars.km $^{-1}$  ( $R^2 = 0.91$ ).

*Data about the GDP of different countries were obtained from the World Bank.*

on Fig. 62 and the data agree well ( $R^2 = 0.91$ ) with a linear dependence between  $L$  and  $G$ . Again, the dispersion indicates that the linear trend should only be understood as an average behaviour and that local particularities can have a strong impact on the important deviations observed. For instance, the United Arab Emirates are far from the average behaviour, with a 52 km network and a GDP of roughly  $3 \cdot 10^5$  million dollars. Yet, the construction of a 1,200 km railway network has been decided in 2010, which would bring the country closer to the average behaviour.

#### 15.4 SUMMARY

We have proposed a general framework to connect the properties of railway and subway systems (ridership, total length and number of stations) to the socio-economic and spatial characteristics of the country or city they are built in (population, area, GDP). Despite their simplicity, our arguments agree satisfactorily with the data we gathered for more than 100 subway systems and 50 railway networks across the world. It should be noted that the noise associated with these data (and sometimes their definition, see Material and Methods) makes it difficult to infer behaviours from the empirical analysis alone. Therefore, the most appropriate way to proceed, we believe, is to make assumptions about the systems and build a model whose predictions can then be tested against data.

This study suggests that the fundamental difference between railways and subways comes from the determination of the interstation distance. While it is imposed by human constraints in the subway case, the railway network has to adapt to the spatial distribution of cities in a country. This remark is at the heart of the different behaviors observed for railways and subways (see Table 11 for a summary of these differences).

The previous arguments are able to explain the average behaviour of various quantities. Nevertheless, it would be interesting to identify

	Subway	Train
$L/N_s$	cste.	$\sqrt{\frac{A}{N_s}}$
$R$	$\frac{P}{A} N_s$	$N_s$
$G$	$N_s$	$L$

Table 11: **Summary of the differences between subways and railways**

We summarize the difference of behaviour between subways and railways. The scaling of the length  $L$  of the network with the number of stations  $N_s$  reveals the different logics behind the growth of these systems. Another difference lies in the total ridership  $R$ : while it depends on the population density  $P/A$  for subways, it only depends on the number of stations  $N_s$  for train networks. Finally, the size of both types of network can be expressed as a function of the wealth of the region, represented here by the GDP  $G$ . However, because the interstation length is constant for subways, the size is better expressed in terms of the number of stations  $N_s$ ; in the case of railway networks, the cost of stations are negligible compared to the building cost of lines, and the size is better expressed in terms of the total length  $L$ .

deviations from these behaviours, and see whether they correlate –for instance– with topological properties of the system, as suggested in [67] or other properties of the network and the region. We think that the relations presented here provide nevertheless a simple framework within which local particularities can be discussed and understood. We also think that this framework could be used as a useful null-model to quantify the efficiency of individual transportation networks, and compare them to each other. This would however require more specific data than those that were available to us.

While we have focused on an average, static description of metro systems, we believe that our study provides a better understanding of how these systems interact with the region they serve. This new insight is a necessary step towards a model for the growth of subway systems that takes the characteristics of the city into account. Indeed, although models of network growth exist, the length of networks and nodes at a given time is usually imposed exogeneously, instead of being linked to the socio-economic properties of the substrate. This study provides a simple approach to these complex problems and could help in building more realistic models, with less exogeneous parameters.

It would be interesting to gather data about the exact structure of all the studied network, so as to study whether there is a relationship between the topology (degree distribution, detour index, etc.) of these networks and properties of the substrate, as was done for the road network in [137].

Finally, gathering historical data should allow to address the problem of the conditions for the appearance of a subway in a city. In particular, we observe empirically that the GDP of the cities that have a subway

system is always larger than about  $10^{10}$  dollars, a fact that calls for a theoretical explanation.



Part VI

CONCLUSION

Self-explanatory title.



## CONCLUSION

*If people never did silly things  
nothing intelligent would ever get done.*

— Ludwig Wittgenstein [142]

In this thesis, we have adopted a ‘physicist’ approach to the study of a system that traditionally belonged to the realm of social sciences: the city. We have tried to show that simple approaches allow to better understand these complex systems. Although simple models with a few variables cannot reproduce all the properties and behaviours of the observed phenomena, they allow us to uncover the dominant mechanisms that are responsible for their most salient features. Does it mean that our approach is the only valid approach? Probably not. Is it useful? Certainly, as it structures our knowledge and sets a solid basis for future investigations.

In the first part, we have reviewed the evolution of the concept of polycentricity in the literature, and the methods used to identify and count the number of centers. Doing so, we provided evidence for the increasing number of activity centers with population size, a phenomenon we called ‘polycentric transition’. We then proposed an out-of-equilibrium, stochastic model of city growth that reproduces the empirical regularity, and explains the transition with the increasing levels of congestion as cities get larger. This model is a substantial improvement over the models presented in the Economics literature: it makes predictions that are supported by data, and allows to identify the mechanisms responsible for the observed phenomena.

In the second part, we further use the model to give a prediction for the scaling exponent of the total distance commuted daily, the total length of the road network, the total delay due to congestion, the quantity of CO<sub>2</sub> emitted, and the surface area with the population size of cities. We successfully test these predictions with data gathered for US urban areas.

In a third part, we focus on the quantitative description of the patterns of residential segregation. For the first time in the quantitative literature, we propose an explicit definition of segregation as a deviation from a random distribution of individuals across the urban space. This definition provides a unifying theoretical framework in which segregation can be empirically characterised. We propose a measure of interaction between the different categories. Building on the information about the attraction and repulsion between categories, we are further

able to propose a definition of classes that is quantitative and unambiguous. The framework also allows us to identify the neighbourhoods where the different classes concentrate, and characterise their properties and spatial arrangement. Finally, we revisit the traditional dichotomy between poor city centers and rich suburbs and provide a measure that is adapted to anisotropic, polycentric cities.

In the fourth and last part, we briefly reviewed the results we have obtained in the study of spatial networks. We first presented a quantitative method to classify cities based on their street patterns, which we applied to a set of 131 cities across the world. Then, we introduced an iterative model for the growth of spatial networks that is based on cost-benefit considerations. The model exhibits interesting features: a crossover between the Minimum Spanning Tree and the star graph, with an intermediate regime characterised by the emergence of spatial hierarchy. Finally, we proposed a general coarse-grained approach – based on a cost-benefit analysis – that accounts for the scaling properties of the main quantities characterizing railway and subway networks (the number of stations, the total length, and the ridership) with the substrate’s population, area and wealth. We showed that the length, number of stations and ridership of subways and rail networks can be estimated knowing the area, population and wealth of the underlying region. These predictions are in good agreement with data gathered for about 140 subway systems and more than 50 railway networks in the world.

The field is still in its infancy compared to more mature sciences, but there are very good reasons to hope for the convergence of knowledge and methods into a new discipline. Into what we may call – following Michael Batty – a Science of Cities [27, 28]. It is difficult at this stage to say what this Science will look like, and what kind of results it can pretend to achieve. Nevertheless, it is tempting to compare the current state of the field to the study of planetary motions before Isaac Newton’s *Philosophiæ Naturalis Principia Mathematica*, or the study of electromagnetism before James Clerk Maxwell’s *A Dynamical Theory of the Electromagnetic Field*; a set of stylized facts and empirical laws that are yet to be unified in a coherent theory.

This is not to say that one should look for a unifying set of equations, or that laws about urban system will have the same permanence as those describing natural phenomena. No two theories are alike – even in Physics. But we believe that the underlying methodological principles have a universal character. Nothing can go fundamentally wrong if data are the ultimate judge of the validity of our theoretical endeavours.

## 16.1 LESSONS LEARNED

The last 3 years have taught me lessons that go beyond simple scientific knowledge.

16.1.1 *Thinking the city*

A first lesson, painstakingly learned during this thesis is that *thinking* the city is as important as *measuring* the city, or *modeling* the city. Concepts guide us and tell us what to measure, what to model. In the same way measures and model can tell us what to think. It would be very naive to believe that scientific enquiries are fueled by the sole discussion between measures and models. In fact, many studies are based upon an hypothesis, a pattern that the author has seen and whose existence she is trying to prove on a quantitative basis.

It is also certainly true that the most difficult and important problems are conceptual in nature. It is impossible to define a city quantitatively before you have formed—with words, possibly drawings—a conceptual picture of what a city is. It is impossible to study segregation before you have logically clarified what one means by segregation. However quantitative, an investigation built upon weak conceptual foundations is unlikely to go anywhere, or to say anything substantial. On the other hand, when the thoughts have settled and the question is clear, one can quickly make a substantial contribution. In this sense, qualitative and quantitative investigations are not incompatible: they are really two sides of the same coin.

16.1.2 *Disciplinary borders*

The topics I had the chance to tackle during these 3 years of PhD were very diverse. In retrospect, this was a real chance. This pushed me to browse a wide literature that encompassed many different disciplines. What I found striking while perusing articles and books is the tendency of the different communities to ignore one another.

The problem, however, is not to blame on individuals. While there may be deliberate omissions here and there, authors are generally willing to cite the appropriate literature when they are aware of its existence. The issue, I believe, is institutional. It stems from the academic organisation of Science, and the existence of disciplinary borders.

But do disciplinary borders still mean anything? While there is an undeniable historical justification to the existence of disciplines, do they still make sense, scientifically speaking? Should the path-dependency in the evolution of the man-made, academic classification of sciences dictate what research avenues are worth being pursued today? At a time when some topics – including cities – get an increasingly multi-

disciplinary attention, these questions are worth asking. Science is fueled by ignorance and questions, not knowledge. It may therefore be time to organise communities around common questions, rather than (overlapping) corpora of knowledge.

## 16.2 IF I HAD TO WRITE A SECOND THESIS (FUTURE DIRECTIONS)

What would I write about – or at least try to – if I had to start my thesis all over again? This is another way of saying: what are the next steps? Many clues can be found in the various parts of this manuscript. Indeed, I have tried to explicit the limitations of the empirical methods and models presented. In these remarks lie many potential avenues for future research. In the following, I will present some other ideas that sprung over the last 3 years.

I would probably start with the basics, with the single noun that was most often printed in these pages: Cities <sup>1</sup>. It is indeed uncomfortable – to say the least – that our most fundamental object, the city, is ill-defined, and that most empirical studies possibly rely on a definition that is not suited to the investigation they undertake. This lack of serious definition compromises the comparison between cities of different countries, or at different points in time. I am, of course, not the first person to acknowledge this empirical shortcoming. In fact, it is a long-lasting worry of geographers who have been trying to produce harmonised database for many years [188]. Yet, we still lack of an unambiguous, theoretically grounded definition of what a city is. And this is problematic, since statistical institutes' results are based on what is believed to be the best definition of the city at a time. Which in turn influences the research on cities. If we want to exhibit robust empirical results, compare the results obtained in different countries, we therefore need to start worrying about the definition of the system we are studying. We need to know *what* cities we are talking about.

Once the boundaries are defined, we can start studying the way objects are scattered within them. By objects, I mean buildings, roads, and first and foremost people. The way we traditionally study the repartition of objects in space is through the study of densities. But density profiles are too complicated to comprehend for our brains, especially when cities get large. So complicated, that an entire sub-field is dedicated to their study: urban form [222, 205, 131]. Authors attempt to solve this problem by providing simple measures that extract a single number from the profile. A single number is however too simple to be able to describe accurately complex spatial distributions. What we need is a meso-scale representation, somewhere between the micro-scale pic-

---

<sup>1</sup> Not verified on data.

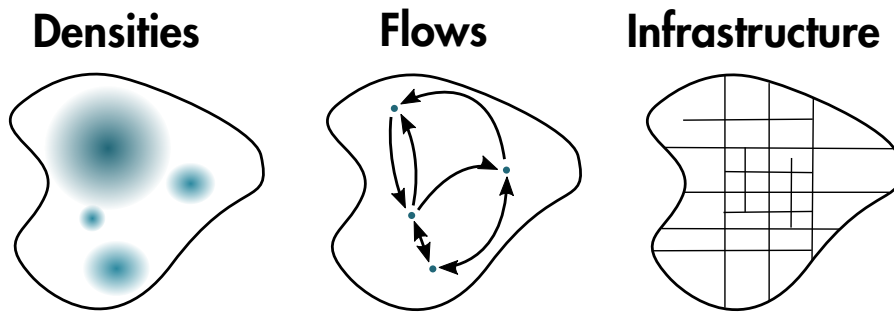


Figure 63: **Intra-urban organisation.** Cities are first and foremost defined by the concentration of populations and various activities. The fact that residences and activities have different locations is responsible for the existence of flows of people, goods, etc. across the urban space. These flows occur on appropriate infrastructure.

ture (the density profile itself) and the macro-scale picture (a single number to summarize the density profile). Hopefully, because ‘centers’ are themselves a mesoscopic structure, their definition should emerge naturally from such a representation.

Once one is able to provide an accurate description of density profiles, the possibilities start to diverge. An obvious worry, when one has a picture of the city’s population at different times of the day, is the way these profile transform one into another. This is linked to commuting—but not only, commuting representing only 20% of total travels in the US [201]—and the study of congestion of networks.

We could first try to explicit the link between the urban form (typically the residential and employment densities) and mobility patterns [143, 57]. For instance, we could wonder: what proportion of commuting flows is due to the spatial mismatch between jobs and residences?

A further worry linked to commuting is that of congestion: understanding how traffic jams are formed, how they propagate and devise strategies to mitigate them, either by influencing the transportation infrastructure, the spatial repartition of residences and employment, or the behaviour of people themselves. This is far from being a recent worry, but there is room for new approaches that leverage the knowledge we have about network and phase transition in physics. A first step in this direction has been made by the authors of [139], but there is surely more to be understood and discovered.

Modeling congestion also implies understanding the individual behaviour of people when they are moving from a point to another in cities. Although most research nowadays assume that people choose the shortest (time or distance) path, GPS data now provide overwhelming evidence that this is not the case [145]. So, while there is a clear need to understand the mesoscopic picture (how congestion spread), there is also a critical need to understand the microscopic picture (how people

behave).

So far we have talked about the movement induced by the spatial mismatch between residential areas and activity areas. One might also want to study the characteristics of the spatial repartition of people. Inhabitants of cities are not just a combination of a latitude and a longitude, a point on a map. Like you and me, they are characterised by different qualities, some of which are measurable: their income, their education level, their ethnicity, etc. A natural question, that has interested sociologists and geographers, is to wonder whether people's residence is independent of these characteristics, or whether these characteristics have an influence on the spatial repartition of individuals.

In this thesis, we provided a rigorous method to study the patterns of segregation in the presence of multiple income categories. The method is far more general, however. It could be used to study the concentration of any category (be it ethnic categories, or certain business types, etc.) in certain regions of the urban space, and quantify the resulting spatial pattern. As a matter of fact, more work is needed to be able to identify the topology and geometry of these distributions. The problem is very close to the description of density pattern described above.

The definition of neighbourhoods (again, a mesoscopic structure) is also not completely satisfactory. Often, it relies on non-overlapping census boundaries that were drawn to maximise the intra-neighbourhood homogeneity and maximise the inter-neighbourhood heterogeneity. Although this may be useful for political institutions to target the most segregated regions of the city, this does not account for how segregation is witnessed by individuals, at an individual level. This has recently been questioned in the Sociology literature, and there has recently been new attempts to define neighbourhoods based on social ties [111].

There are many more ideas that would deserve to be explored, many more topics that are worthy of attention. I hope the years to come will give me the opportunity to address some of them. But not now; this thesis has to stop somewhere.



- [10] Henry Adams. *The Education of Henry Adams*. Vintage, 1990.
- [11] David J. Aldous and Julian Shun. Connected spatial networks over random points and a route-length statistic. *Statistical Science*, 25(3):275–288, 2010.
- [12] William Alonso. *Location and land use: Toward a general theory of land rent*. Harvard University Press, 1964.
- [13] Alex Anas, Richard Arnott, and Kenneth A. Small. Urban spatial structure. *Journal of Economic Literature*, 36(3):1426–1464, 1998.
- [14] Chris Anderson. The end of theory: The data deluge makes the scientific method obsolete, 2008. URL [http://archive.wired.com/science/discoveries/magazine/16-07/pb\\_theory/](http://archive.wired.com/science/discoveries/magazine/16-07/pb_theory/). [Online; Posted on 23/06/2008].
- [15] Philippe Apparicio. Les indices de ségrégation résidentielle: un outil intégré dans un système d'information géographique. *Cybergeo: European Journal of Geography*, 2000.
- [16] Elsa Arcaute, Erez Hatna, Peter Ferguson, Hyejin Youn, Anders Johansson, and Michael Batty. Constructing cities, deconstructing scaling laws. *Journal of The Royal Society Interface*, 12(102):20140745, 2014.
- [17] Felix Auerbach. Das gesetz der bevölkerungskonzentration. *Petermanns Geographische Mitteilungen*, 59:73–76, 1913.
- [18] Paul Bairoch. *De Jéricho à Mexico: villes et économie dans l'histoire*. Gallimard, 1985.
- [19] Bela Balassa. Trade liberalisation and "revealed" comparative advantage<sup>1</sup>. *The Manchester School*, 33(2):99–123, 1965.
- [20] Jayanth R. Banavar, Amos Maritan, and Andrea Rinaldo. Size and form in efficient transportation networks. *Nature*, 399:130–132, 1999.
- [21] Grigory I. Barenblatt. *Scaling, self-similarity, and intermediate asymptotics*. Cambridge University Press, 1996.
- [22] Marc Barthelemy. Spatial networks. *Physics Reports*, 499(1):1–101, 2011.
- [23] Marc Barthelemy and Alessandro Flammini. Modeling urban street patterns. *Physical Review Letters*, 100(13):138702, 2008.
- [24] Marc Barthelemy, Patricia Bordin, Henri Berestycki, and Maurizio Gribaudi. Self-organization versus top-down planning in the evolution of a city. *Scientific Reports*, 3:2153, 2013.

- [25] Michael Batty. New ways of looking at cities. *Nature*, 377:574, 1995.
- [26] Michael Batty. The size, scale, and shape of cities. *Science*, 319:769–771, 2008.
- [27] Michael Batty. Building a science of cities. *Cities*, 29:S9–S16, 2012.
- [28] Michael Batty. *The new science of cities*. The MIT Press, 2013.
- [29] Michael Batty and Peter Ferguson. Defining city size. *Environment and Planning B: Planning and Design*, 38(5):753–756, 2011.
- [30] Wendell Bell. A probability model for the measurement of ecological segregation. *Social Forces*, 32(4):357–364, 1954.
- [31] Isaiah Berlin. Historical inevitability. In Henry Hardy and Roger Hausheer, editors, *The proper study of mankind*. Vintage, 2013.
- [32] Jonathan A. Bernstein, Neil Alexis, Charles Barnes, I. Leonard Bernstein, Andre Nel, David Peden, David Diaz-Sanchez, Susan M. Tarlo, and P. Brock Williams. Health effects of air pollution. *Journal of Allergy and Clinical Immunology*, 114(5):1116–1123, 2004.
- [33] Sandrine Berroir, Hélène Mathian, Thérèse Saint-Julien, and Lena Sanders. La mobilité dans la construction du polycentrisme métropolitain. In M. Thériault and F. Des Rosiers, editors, *Information géographique et dynamiques urbaines*, pages 10–100. Hermès, 2008.
- [34] Brian J.L. Berry. Cities as systems within systems of cities. *Papers in Regional Science*, 13(1):147–163, 1964.
- [35] Brian J.L. Berry. *Geography of market centers and retail distribution*. Englewood Cliffs, 1967.
- [36] Brian J.L. Berry. Geography’s quantitative revolution: Initial conditions, 1954-1960. A personal memoir. *Urban Geography*, 14(5):434–441, 1993.
- [37] Luís Bettencourt and Geoffrey West. A unified theory of urban living. *Nature*, 467:912–913, 2010.
- [38] Luís M.A. Bettencourt. The origins of scaling in cities. *Science*, 340:1438–1441, 2013.
- [39] Luís M.A. Bettencourt, José Lobo, Dirk Helbing, Christian Kühnert, and Geoffrey B. West. Growth, innovation, scaling, and the pace of life in cities. *Proceedings of the National Academy of Sciences, U.S.A.*, 104(17):7301–7306, 2007.

- [40] Luis M.A. Bettencourt, José Lobo, Deborah Strumsky, and Geoffrey B. West. Urban scaling and its deviations: Revealing the structure of wealth, innovation and crime across cities. *PLoS ONE*, 5(11):e13541, 2010.
- [41] Luís M.A. Bettencourt, Horacio Samaniego, and Hyejin Youn. Professional diversity and the productivity of cities. *Scientific Reports*, 4:5393, 2014.
- [42] William R. Black. An iterative model for generating transportation networks. *Geographical Analysis*, 3(3):283–288, 1971.
- [43] Vincent D. Blondel, Markus Esch, Connie Chan, Fabrice Clérot, Pierre Deville, Etienne Huens, Frédéric Morlot, Zbigniew Smoreda, and Cezary Ziemlicki. Data for development: the D4D challenge on mobile phone data. *arXiv:1210.0137*, 2012.
- [44] Jean-Philippe Bouchaud. Economics needs a scientific revolution. *Nature*, 455:1181–1181, 2008.
- [45] Jérémie Bouttier, Philippe Di Francesco, and Emmanuel Guitter. Planar maps as labeled mobiles. *Electronic Journal of Combinatorics*, 11(1):R69, 2004.
- [46] Ronald R. Boyce. Changing patterns of urban land consumption. *The Professional Geographer*, 15(2):19–24, 1963.
- [47] David Branston. Link capacity functions: A review. *Transportation Research*, 10(4):223–236, 1976.
- [48] Anne Bretagnolle and François Delisle. Formes de villes en Europe et aux Etats-Unis. *Mappemonde*, 97, 2010.
- [49] Jan K. Brueckner. The structure of urban equilibria: A unified treatment of the Muth-Mills model. *Handbook of Regional and Urban Economics*, 2:821–845, 1987.
- [50] Jan K. Brueckner, Jacques-François Thisse, and Yves Zenou. Why is central Paris rich and downtown Detroit poor?: An amenity-based theory. *European Economic Review*, 43(1):91–107, 1999.
- [51] Chris Brunsdon, Stewart Fotheringham, and Martin Charlton. Geographically weighted regression. *Journal of the Royal Statistical Society: Series D (The Statistician)*, 47(3):431–443, 1998.
- [52] William Bunge. *Theoretical Geography*. Royal University of Lund, 1962.
- [53] Alessio Cardillo, Salvatore Scellato, Vito Latora, and Sergio Porta. Structural properties of planar graphs of urban street patterns. *Physical Review E*, 73(6):066107, 2006.

- [54] Jean-Claude Chamboredon and Madeleine Lemaire. Proximité spatiale et distance sociale. Les grands ensembles et leur peuplement. *Revue française de sociologie*, pages 3–33, 1970.
- [55] Sonic H.Y. Chan, Reik V. Donner, and Stefan Lämmer. Urban road networks: spatial networks with universal geometric features? *The European Physical Journal B*, 84(4):563–577, 2011.
- [56] Richard J. Chorley and Peter Haggett. *Socio-economic Models in Geography*. Methuen, 1968.
- [57] Tufayel A. Chowdhury, Darren M. Scott, and Pavlos S. Kanaroglou. Urban form and commuting efficiency: a comparative analysis across time and space. *Urban Studies*, 50(1):191–207, 2013.
- [58] Walter Christaller. *Die zentralen orte in Süddeutschland*. Gustav Fischer, 1933.
- [59] Colin Clark. Urban population densities. *Journal of the Royal Statistical Society. Series A (General)*, 114(4):490–496, 1951.
- [60] Aaron Clauset, Cosma R. Shalizi, and Mark E.J. Newman. Power-law distributions in empirical data. *SIAM review*, 51(4):661–703, 2009.
- [61] T.M. Coburn. The London-Birmingham motorway: Traffic and economics, 1960.
- [62] Clémentine Cottineau, Erez Hatna, Elsa Arcaute, and Michael Batty. Paradoxical interpretations of urban scaling laws. *ArXiv:1507.07878*, 2015.
- [63] Thomas Courtat, Catherine Gloaguen, and Stephane Douady. Mathematics and morphogenesis of cities: A geometrical approach. *Physical Review E*, 83(3):036106, 2011.
- [64] Matthieu Cristelli, Michael Batty, and Luciano Pietronero. There is more than a power law in zipf. *Scientific Reports*, 2:812, 2012.
- [65] Paolo Crucitti, Vito Latora, and Sergio Porta. Centrality measures in spatial networks of urban streets. *Physical Review E*, 73(3):036125, 2006.
- [66] Daniel C. Dennett. Real patterns. *The Journal of Philosophy*, pages 27–51, 1991.
- [67] Sybil Derrible and Christopher Kennedy. Network analysis of world subway systems using updated graph theory. *Transportation Research Record: Journal of the Transportation Research Board*, (2112):17–25, 2009.

- [68] Pierre Deville, Catherine Linard, Samuel Martin, Marius Gilbert, Forrest R. Stevens, Andrea E. Gaughan, Vincent D. Blondel, and Andrew J. Tatem. Dynamic population mapping using mobile phone data. *Proceedings of the National Academy of Sciences, U.S.A.*, 111(45):15888–15893, 2014.
- [69] Philip M. Dixon, Jacob Weiner, Thomas Mitchell-Olds, and Robert Woodley. Bootstrapping the Gini coefficient of inequality. *Ecology*, pages 1548–1551, 1987.
- [70] Vedia Dökmeci and Lale Berköz. Transformation of istanbul from a monocentric to a polycentric city. *European Planning Studies*, 2(2):193–205, 1994.
- [71] Pierre Duhem. *La théorie physique: son objet, sa structure*. Vrin, 1997.
- [72] Otis D. Duncan and Beverly Duncan. A methodological analysis of segregation indexes. *American Sociological Review*, 20(2):210–217, 1955.
- [73] Gabriel Dupuy and Lucien G. Benguigui. Sciences urbaines: interdisciplinarités passive, naïve, transitive, offensive. *Métropoles*, (16), 2015.
- [74] Gilles Duranton and Henry G. Overman. Testing for localization using micro-geographic data. *The Review of Economic Studies*, 72(4):1077–1106, 2005.
- [75] Gilles Duranton and Diego Puga. Micro-foundations of urban agglomeration economies. *Handbook of Regional and Urban Economics*, 4:2063–2117, 2004.
- [76] Freeman Dyson. A meeting with enrico fermi. *Nature*, 427:297–297, 2004.
- [77] Freeman J. Dyson. Statistical theory of the energy levels of complex systems. I. *Journal of Mathematical Physics*, 3(1):140–156, 1962.
- [78] Nathan Eagle, Michael Macy, and Rob Claxton. Network diversity and economic development. *Science*, 328:1029–1031, 2010.
- [79] Albert Einstein, Boris Podolsky, and Nathan Rosen. Can quantum-mechanical description of physical reality be considered complete? *Physical Review*, 47(10):777–780, 1935.
- [80] Mustafa Emirbayer. Manifesto for a relational sociology. *American Journal of Sociology*, 103(2):281–317, 1997.
- [81] Sven Erlander and Neil F. Stewart. *The gravity model in transportation analysis: theory and extensions*. VSP, 1990.

- [82] Alex Fabrikant, Elias Koutsoupias, and Christos H. Papadimitriou. Heuristically optimized trade-offs: A new paradigm for power laws in the internet. In Peter Widmayer, Stephan Eidenbenz, Francisco Triguero, Rafael Morales, Ricardo Conejo, and Matthew Hennessy, editors, *Automata, Languages and Programming*, volume 2380 of *Lecture Notes in Computer Science*, pages 110–122. Springer, 2002.
- [83] William Feller. *An introduction to probability theory and its applications*. John Wiley & Sons, 1950.
- [84] Julie Fen-Chong. *Organisation spatio-temporelle des mobilités révélées par la téléphonie mobile en Ile-de-France*. PhD thesis, Université Panthéon-Sorbonne-Paris I, 2012.
- [85] Paul Feyerabend. *Against Method*. New Left Books, 1975.
- [86] Richard P. Feynman. *The character of physical law*. MIT Press, 1965.
- [87] Marcin Fialkowski and Agnieszka Bitner. Universal rules for fragmentation of land by humans. *Landscape Ecology*, 23(9):1013–1022, 2008.
- [88] Michail Fragkias, José Lobo, Deborah Strumsky, and Karen C. Seto. Does size matter? Scaling of  $CO_2$  emissions and U.S. urban areas. *PLoS One*, 8(6):e64727, 2013.
- [89] Gerald F. Frasco, Jie Sun, Hernán D. Rozenfeld, and Daniel Ben-Avraham. Spatially distributed social complex networks. *Physical Review X*, 4(1):011008, 2014.
- [90] Masahisa Fujita. *Urban economic theory: land use and city size*. Cambridge University Press, 1989.
- [91] Masahisa Fujita and Hideaki Ogawa. Multiple equilibria and structural transition of non-monocentric urban configurations. *Regional Science and Urban Economics*, 12(2):161–196, 1982.
- [92] Masahisa Fujita and Jacques-François Thisse. *Economics of agglomeration: Cities, industrial location, and globalization*. Cambridge University Press, 2013.
- [93] Masahisa Fujita, Paul R. Krugman, and Anthony J. Venables. *The spatial economy: Cities, regions, and international trade*. The MIT Press, 2001.
- [94] Richard A. Fuller and Kevin J. Gaston. The scaling of green space coverage in European cities. *Biology Letters*, 5(3):352–355, 2009.
- [95] Joel Garreau. *Edge city: Life on the new frontier*. Doubleday, New-York, 1991.

- [96] Michael T. Gastner and Mark E.J. Newman. Shape and efficiency in spatial distribution networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2006(01):P01015, 2006.
- [97] Laetitia Gauvin, Annick Vignes, and Jean-Pierre Nadal. Modeling urban housing market dynamics: can the socio-spatial segregation preserve some social diversity? *Journal of Economic Dynamics and Control*, 37(7):1300–1321, 2013.
- [98] Genevieve Giuliano and Kenneth A. Small. Subcenters in the Los Angeles region. *Regional Science and Urban Economics*, 21(2):163–182, 1991.
- [99] Edward L. Glaeser and Matthew E. Kahn. The greenness of cities: carbon dioxide emissions and urban development. *Journal of Urban Economics*, 67(3):404–418, 2010.
- [100] Edward L. Glaeser, Matthew E. Kahn, and Jordan Rappaport. Why do the poor live in cities? The role of public transportation. *Journal of Urban Economics*, 63(1):1–24, 2008.
- [101] Leon Glass and Waldo R. Tobler. Uniform distribution of objects in a homogeneous field: cities on a plain. *Nature*, 233:67–68, 1971.
- [102] Marta C. Gonzalez, Cesar A. Hidalgo, and Albert-Laszlo Barabasi. Understanding individual human mobility patterns. *Nature*, 453(7196):779–782, 2008.
- [103] Peter Gordon and Harry W. Richardson. Beyond polycentricity: the dispersed metropolis, Los Angeles, 1970-1990. *Journal of the American Planning Association*, 62(3):289–295, 1996.
- [104] Daniel A. Griffith. Evaluating the transformation from a monocentric to a polycentric city. *The Professional Geographer*, 33(2):189–196, 1981.
- [105] Marianne Guérois. *Les formes des villes européennes vues du ciel*. PhD thesis, UMR Géographie-Cités, 2003.
- [106] Peter Haggett. *Locational Analysis in Humans Geography*. Edward Arnold, 1966.
- [107] Peter Haggett and Richard J. Chorley. *Network Analysis in Geography*. Edward Arnold, 1969.
- [108] David Harvey. *Explanation in Geography*. Edward Arnold, 1969.
- [109] Robert Hass. *Sun Under Wood*. Harper Collins, 1998.
- [110] Bill Hillier and Julienne Hanson. *The social logic of space*. Cambridge University Press, 1984.

- [111] John R. Hipp, Robert W. Faris, and Adam Boessen. Measuring ‘neighborhood’: Constructing network neighborhoods. *Social Networks*, 34(1):128–140, 2012.
- [112] J.D. Hunter. Matplotlib: A 2D graphics environment. *Computing In Science & Engineering*, 9(3):90–95, 2007.
- [113] Jane Jacobs. *The death and life of great American cities*. Random House, 1961.
- [114] Julius Jahn, Calvin F. Schmid, and Clarence Schrag. The measurement of ecological segregation. *American Sociological Review*, 12(3):293–303, 1947.
- [115] David R. James and Karl E. Taeuber. Measures of segregation. Technical report, University of Wisconsin – Madison, Center for Demography and Ecology, 1982.
- [116] Paul A. Jargowsky. Take the money and run: economic segregation in U.S. metropolitan areas. *American Sociological Review*, 61(6):984–998, 1996.
- [117] Mark Jefferson. The anthropography of some great cities. A study in distribution of population. *Bulletin of the American Geographical Society*, 41(9):537–566, 1909.
- [118] Pablo Jensen. Network-based predictions of retail store commercial categories and optimal locations. *Physical Review E*, 74(3):035101, 2006.
- [119] Bin Jiang. A topological pattern of urban street networks: universality and peculiarity. *Physica A: Statistical Mechanics and its Applications*, 384(2):647–655, 2007.
- [120] Bin Jiang and Christophe Claramunt. Topological analysis of urban street networks. *Environment and Planning B: Planning and Design*, 31(1):151–162, 2004.
- [121] Karel J. Kinsky. *Structure of transportation networks: relationships between network geometry and regional characteristics*. PhD thesis, University of Chicago, 1963.
- [122] Eleni Katifori and Marcelo O. Magnasco. Quantifying loopy network architectures. *PloS ONE*, 7(6):e37994, 2012.
- [123] Leonard Kaufman and Peter J. Rousseeuw. *Finding groups in data: an introduction to cluster analysis*, volume 344. John Wiley & Sons, 2009.
- [124] Peter Kemper and Roger Schmenner. The density gradient for manufacturing industry. *Journal of Urban Economics*, 1(4):410–427, 1974.



- [125] Leslie J. King. *Statistical Analysis in Geography*. Prentice-Hall, 1969.
- [126] Jon M. Kleinberg. Navigation in a small world. *Nature*, 406: 845–845, 2000.
- [127] Michael Kuby, Anthony Barranda, and Christopher Upchurch. Factors influencing light-rail station boardings in the united states. *Transportation Research Part A: Policy and Practice*, 38(3):223–247, 2004.
- [128] James Ladyman, James Lambert, and Karoline Wiesner. What is a complex system? *European Journal for Philosophy of Science*, 3(1):33–67, 2013.
- [129] Stefan Lämmer, Björn Gehlsen, and Dirk Helbing. Scaling laws in the spatial structure of urban road networks. *Physica A: Statistical Mechanics and its Applications*, 363(1):89–95, 2006.
- [130] Florent Le Néchet. *Multi-scalar approach of links between daily mobility, spatial structure and sustainability of European metropolitan areas. Case studies: Paris and Rhine-Ruhr metropolitan regions*. PhD thesis, Université Paris-Est, 2010.
- [131] Florent Le Néchet. De la forme urbaine à la structure métropolitaine: une typologie de la configuration interne des densités pour les principales métropoles européennes de l’Audit Urbain. *Cybergeo: European Journal of Geography*, 2015.
- [132] Maxime Lenormand, Sylvie Huet, Floriana Gargiulo, and Guillaume Deffuant. A universal model of commuting networks. *PLoS ONE*, 7(10):e45985, 10 2012.
- [133] Maxime Lenormand, Miguel Picornell, Oliva G. Cantu-Ros, Antonia Tugores, Thomas Louail, Ricardo Herranz, Marc Barthelemy, Enrique Frias-Martinez, and Jose J. Ramasco. Cross-checking different sources of mobility information. *PLoS ONE*, 9(8):e105184, 2014.
- [134] Maxime Lenormand, Antònia Tugores, Pere Colet, and José J. Ramasco. Tweets on the road. *PLoS ONE*, 9(8):e105407, 2014.
- [135] Maxime Lenormand, Thomas Louail, Oliva G. Cantú-Ros, Miguel Picornell, Ricardo Herranz, Juan M. Arias, Marc Barthelemy, Maxi San Miguel, and José J. Ramasco. Influence of sociodemographics on human mobility. *Scientific Reports*, 5, 2015.
- [136] David Levinson. Density and dispersion: the co-development of land use and rail in london. *Journal of Economic Geography*, 8(1):55–77, 2008.

- [137] David Levinson. Network structure and city size. *PLoS ONE*, 7(1):e29721, 2012.
- [138] David Levinson and Bhanu Yerra. Self-organization of surface transportation networks. *Transportation Science*, 40(2):179–188, 2006.
- [139] Daqing Li, Bowen Fu, Yunpeng Wang, Guangquan Lu, Yehiel Berezin, H. Eugene Stanley, and Shlomo Havlin. Percolation transition in dynamical traffic network with evolving critical bottlenecks. *Proceedings of the National Academy of Sciences, U.S.A.*, 112(3):669–672, 2015.
- [140] Peter Lobmayer and Richard G. Wilkinson. Inequality, residential segregation by income, and mortality in U.S. cities. *Journal of Epidemiology and Community Health*, 56(3):183–187, 2002.
- [141] Thomas Louail, Maxime Lenormand, Oliva G.C. Ros, Miguel Picornell, Ricardo Herranz, Enrique Frias-Martinez, José J. Ramasco, and Marc Barthelemy. From mobile phone data to the spatial structure of cities. *Scientific Reports*, 4:5276, 2014.
- [142] C. Grant Luckhardt. *Wittgenstein, sources and perspectives*. St Augustine’s Press, 1979.
- [143] Kang-Rae Ma and David Banister. Excess commuting: a critical review. *Transport Reviews*, 26(6):749–767, 2006.
- [144] Hernán A. Makse, Shlomo Havlin, and H. Eugene Stanley. Modelling urban growth patterns. *Nature*, 377:608–612, 1995.
- [145] Ed J. Manley, JD Addison, and T Cheng. Shortest path or anchor-based route choice: a large-scale empirical analysis of minicab routing in london. *Journal of Transport Geography*, 43:123–139, 2015.
- [146] Eric Marcon and Florence Puech. Measures of the geographic concentration of industries: improving distance-based methods. *Journal of Economic Geography*, page lbp056, 2009.
- [147] Stephen Marshall. *Streets and patterns*. Routledge, 2004.
- [148] Matteo Marsili and Yi-Cheng Zhang. Interacting individuals leading to Zipf’s law. *Physical Review Letters*, 80(12):2741, 1998.
- [149] Douglas S. Massey and Nancy A. Denton. The dimensions of residential segregation. *Social Forces*, 67(2):281–315, 1988.
- [150] Douglas S. Massey and Nancy A. Denton. *American apartheid: Segregation and the making of the underclass*. Harvard University Press, 1993.

- [151] Frank J. Massey Jr. The Kolmogorov-Smirnov test for goodness of fit. *Journal of the American statistical Association*, 46(253): 68–78, 1951.
- [152] A. Paolo Masucci, Duncan Smith, Andrew Crooks, and Michael Batty. Random planar graphs and the London street network. *The European Physical Journal B*, 71(2):259–271, 2009.
- [153] A. Paolo Masucci, Elsa Arcaute, Jiaqiu Wang, Erez Hatna, Kiril Stanilov, and Michael Batty. Logistic growth and ergodic properties of urban forms. *arXiv:1504.07380*, 2015.
- [154] Ryoji Matsunaka, Tetsuharu Oba, Dai Nakagawa, Motoya Nagao, and Justin Nawrocki. International comparison of the relationship between urban structure and the service level of urban public transportation—a comprehensive analysis in local cities in Japan, France and Germany. *Transport Policy*, 30:26–39, 2013.
- [155] John F. McDonald. The identification of urban employment subcenters. *Journal of Urban Economics*, 21(2):242–258, 1987.
- [156] Daniel P. McMillen. Nonparametric employment subcenter identification. *Journal of Urban Economics*, 50(3):448–473, 2001.
- [157] Daniel P. McMillen and Stefani C. Smith. The number of subcenters in large urban areas. *Journal of Urban Economics*, 53(3): 321–338, 2003.
- [158] Patricia C. Melo, Daniel J. Graham, and Robert B. Noland. A meta-analysis of estimates of urban agglomeration economies. *Regional Science and Urban Economics*, 39(3):332–342, 2009.
- [159] Paul Meuriot. *Des agglomérations urbaines dans l’Europe contemporaine: essai sur les causes, les conditions, les conséquences de leur développement*. Belin frères, 1898.
- [160] Yuriy Mileyko, Herbert Edelsbrunner, Charles A. Price, and Joshua S. Weitz. Hierarchical ordering of reticular networks. *PLoS ONE*, 7(6):e36715, 2012.
- [161] Edwin S. Mills. An aggregative model of resource allocation in a metropolitan area. *The American Economic Review*, 57(2):197–210, 1967.
- [162] Edwin S. Mills. *Studies in the Structure of the Urban Economy*. The Johns Hopkins Press, 1972.
- [163] Lewis Mumford. *The city in history. its origins, its transformation, and its prospects*. Harcourt, Brace & World, 1961.
- [164] Richard Muth. *Cities and housing: The spatial patterns of urban residential land use*. University of Chicago Press, 1969.

- [165] Peter W.G. Newman and Jeffrey R. Kenworthy. Gasoline consumption and cities: a comparison of US cities with a global survey. *Journal of the american planning association*, 55(1):24–37, 1989.
- [166] Stig Nordbeck. The law of allometric growth, 1965.
- [167] Anastasios Noulas, Salvatore Scellato, Renaud Lambiotte, Massimiliano Pontil, and Cecilia Mascolo. A tale of many cities: universal patterns in human urban mobility. *PLoS ONE*, 7(5):e37027, 2012.
- [168] John Odland. The conditions for multi-center cities. *Economic Geography*, 54(3):234–244, 1978.
- [169] Morton E. O’Kelly. A geographer’s analysis of hub-and-spoke networks. *Journal of Transport Geography*, 6(3):171–186, 1998.
- [170] Morton E. O’Kelly, Deborah Bryan, Darko Skorin-Kapov, and Jadranka Skorin-Kapov. Hub network design with single and multiple allocation: A computational study. *Location Science*, 4(3):125–138, 1996.
- [171] Erneson A. Oliveira, José S. Andrade Jr., and Hernán A. Makse. Large cities are less green. *Scientific Reports*, 4:4235, 2014.
- [172] Stan Openshaw. The modifiable areal unit problem, 1984.
- [173] J. Robert Oppenheimer. *Science and common understanding*. Simon and Schuster, 1954.
- [174] Naomi Oreskes. The scientific consensus on climate change. *Science*, 306:1686–1686, 2004.
- [175] Giorgio Parisi. Complex systems: a physicist’s viewpoint. *Physica A: Statistical Mechanics and its Applications*, 263(1):557–564, 1999.
- [176] Fabien Paulus. *Coévolution dans les systèmes de villes: croissance et spécialisation des aires urbaines françaises de 1950 à 2000*. PhD thesis, Université Panthéon-Sorbonne-Paris I, 2004.
- [177] Alan Penn. Space syntax and spatial cognition or why the axial line? *Environment and Behavior*, 35(1):30–65, 2003.
- [178] Marko Popović, Hrvoje Štefančić, and Vinko Zlatić. Geometric origin of scaling in large traffic networks. *Physical Review Letters*, 109(20):208701, 2012.
- [179] Sergio Porta, Paolo Crucitti, and Vito Latora. The network analysis of urban streets: a primal approach. *Environment and Planning B: Planning and Design*, 33(5):705, 2006.

- [180] Sergio Porta, Paolo Crucitti, and Vito Latora. The network analysis of urban streets: a dual approach. *Physica A: Statistical Mechanics and its Applications*, 369(2):853–866, 2006.
- [181] Sergio Porta, Ombretta Romice, J. Alexander Maxwell, Peter Russell, and Darren Baird. Alterations in scale: patterns of change in main street networks across time and space. *Urban Studies*, 51(16):3383–3400, 2014.
- [182] Robert C. Prim. Shortest connection networks and some generalizations. *Bell System Technical Journal*, 36(6):1389–1401, 1957.
- [183] Denise Pumain. Scaling laws and urban systems, 2004.
- [184] Denise Pumain. Une théorie géographique des villes. *Bulletin de la Société géographique de Liège*, (55):5–15, 2010.
- [185] Denise Pumain. Urban systems dynamics, urban growth and scaling laws: The question of ergodicity. In Juval Portugali, Han Meyer, Egbert Stolk, and Ekim Tan, editors, *Complexity Theories of Cities Have Come of Age*, pages 91–103. Springer, 2012.
- [186] Denise Pumain and François Moriconi-Ebrard. City size distributions and metropolisation. *Geojournal*, 43(4):307–314, 1997.
- [187] Denise Pumain, Fabien Paulus, Céline Vacchiani-Marcuzzo, and José Lobo. An evolutionary theory for interpreting urban scaling laws. *Cybergeog: European Journal of Geography*, 2006.
- [188] Denise Pumain, Elfie Swerts, Clémentine Cottineau, Céline Vacchiani-Marcuzzo, Antonio Ignazzi, Anne Bretagnolle, François Delisle, Robin Cura, Liliane Lizzi, and Solène Baffi. Multilevel comparison of large urban systems. *arXiv:1501.06408*, 2015.
- [189] Sean F. Reardon and Glenn Firebaugh. Measures of multigroup segregation. *Sociological Methodology*, 32(1):33–67, 2002.
- [190] Sean F. Reardon and David O’Sullivan. Measures of spatial segregation. *Sociological Methodology*, 34(1):121–162, 2004.
- [191] Christian L. Redfearn. The topography of metropolitan employment: Identifying centers of employment in a polycentric urban area. *Journal of Urban Economics*, 61(3):519–541, 2007.
- [192] William J. Reilly. *The law of retail gravitation*. Knickerbox Press, 1931.
- [193] Brian D. Ripley. *Spatial statistics*. John Wiley & Sons, 1981.
- [194] Martin Rosvall, Ala Trusina, Petter Minnhagen, and Kim Sneppen. Networks and cities: An information perspective. *Physical Review Letters*, 94(2):028701, 2005.

- [195] Camille Roth, Soong Moon Kang, Michael Batty, and Marc Barthélemy. A long-time limit for world subway networks. *Journal of The Royal Society Interface*, 9(75):0259, 2012.
- [196] Hernán D. Rozenfeld, Diego Rybski, José S. Andrade Jr., Michael Batty, H. Eugene Stanley, and Hernán A. Makse. Laws of population growth. *Proceedings of the National Academy of Sciences, U.S.A.*, 105(48):18702–18707, 2008.
- [197] Diego Rybski, Till Sterzel, Dominik E. Reusser, Anna-Lena Winz, Christina Fichtner, and Jürgen P. Kropp. Cities as nuclei of sustainability? *arXiv:1304.4406*, 2013.
- [198] Marta Sales-Pardo, Roger Guimera, André A. Moreira, and Luís A. Nunes Amaral. Extracting the hierarchical organization of complex systems. *Proceedings of the National Academy of Sciences, U.S.A.*, 104(39):15224–15229, 2007.
- [199] Horacio Samaniego and Melanie E. Moses. Cities as organisms: Allometric scaling of urban road networks. *Journal of Transport and Land use*, 1(1), 2008.
- [200] Lena Sanders. Géographie quantitative et analyse spatiale: quelles formes de scientificités? In Thierry Martin, editor, *Les sciences sociales sont-elles des sciences?*, pages 71–91. Vuibert, 2011.
- [201] Adella Santos, Nancy McGuckin, Hikari Y. Nakamoto, Danielle Gray, and Susan Liss. Summary of travel trends: 2009 National Households Travel Survey. Technical Report FHWA-PL-11-022, Federal Highway Administration, Washington, DC, June 2011. URL <http://nhts.ornl.gov/2009/pub/stt.pdf>.
- [202] Thomas C. Schelling. Dynamic models of segregation. *Journal of Mathematical Sociology*, 1(2):143–186, 1971.
- [203] Markus Schläpfer, Luís M.A. Bettencourt, Sébastien Grauwin, Mathias Raschke, Rob Claxton, Zbigniew Smoreda, Geoffrey B. West, and Carlo Ratti. The scaling of human interactions with city size. *Journal of The Royal Society Interface*, 11(98):0789, 2014.
- [204] Markus Schwabe. Residential segregation in the largest French cities (1968-1999): in search of an urban model. *Cybergeo: European Journal of Geography*, 2011.
- [205] Nina Schwarz. Urban form revisited-selecting indicators for characterising European cities. *Landscape and Urban Planning*, 96(1): 29–47, 2010.
- [206] Cosma Rohilla Shalizi. Scaling and hierarchy in urban economies. *arXiv:1102.4101*, 2011.

- [207] Aaron Sim, Sophia N. Yaliraki, Mauricio Barahona, and Michael P.H. Stumpf. Great cities look small. *Journal of The Royal Society Interface*, 12(109):20150315, 2015.
- [208] Filippo Simini, Marta C. González, Amos Maritan, and Albert-László Barabási. A universal model for mobility and migration patterns. *Nature*, 484:96–100, 2012.
- [209] Filippo Simini, Amos Maritan, and Zoltán Nédá. Human mobility in a continuum approach. *PLoS ONE*, 8(3):e60069, 2013.
- [210] Herbert A. Simon. The architecture of complexity. *Proceedings of the American Philosophical Society*, 106(6):467–482, 1962.
- [211] Chaoming Song, Tal Koren, Pu Wang, and Albert-László Barabási. Modelling the scaling properties of human mobility. *Nature Physics*, 6(10):818–823, 2010.
- [212] Chaoming Song, Zehui Qu, Nicholas Blumm, and Albert-László Barabási. Limits of predictability in human mobility. *Science*, 327(5968):1018–1021, 2010.
- [213] Michael Southworth and Eran Ben-Joseph. *Streets and the Shaping of Towns and Cities*. Island Press, 2003.
- [214] John Q. Stewart. Suggested principles of "social physics". *Science*, 106:179–180, 1947.
- [215] John Q. Stewart. Demographic gravitation: evidence and applications. *Sociometry*, pages 31–58, 1948.
- [216] John Q. Stewart and William Warntz. Physics of population distribution. *Journal of Regional Science*, 1(1):99–121, 1958.
- [217] Samuel A. Stouffer. Intervening opportunities: a theory relating mobility and distance. *American Sociological Review*, 5(6):845–867, 1940.
- [218] Emanuele Strano, Vincenzo Nicosia, Vito Latora, Sergio Porta, and Marc Barthélemy. Elementary processes governing the evolution of road networks. *Scientific Reports*, 2:296, 2012.
- [219] Darcy W. Thompson. *On growth and form*. Cambridge University Press, 1942.
- [220] Waldo R. Tobler. Satellite confirmation of settlement size coefficients. *Area*, 1(3):30–34, 1969.
- [221] Waldo R. Tobler. A computer movie simulating urban growth in the Detroit region. *Economic Geography*, 46:234–240, 1970.
- [222] Yu-Hsin Tsai. Quantifying urban form: compactness versus 'sprawl'. *Urban Studies*, 42(1):141–161, 2005.

- [223] Howard Veregin and Waldo R. Tobler. Allometric relationships in the structure of street-level databases. *Computers, Environment and Urban Systems*, 21(3):277–290, 1997.
- [224] Geoffrey B. West, James H. Brown, and Brian J. Enquist. A general model for the origin of allometric scaling laws in biology. *Science*, 276:122–126, 1997.
- [225] Michael J. White. The measurement of spatial segregation. *American Journal of Sociology*, 88(5):1008–1018, 1983.
- [226] Michael J. White. Segregation and diversity measures in population distribution. *Population Index*, 52(2):198–221, 1986.
- [227] Galen Wilkerson, Ramin Khalili, and Stefan Schmid. Urban mobility scaling: Lessons from ‘little data’. In *IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*, pages 777–782, 2014.
- [228] Ludwig Wittgenstein. *Tractatus logico-philosophicus*. Dover, 1998.
- [229] Feng Xie and David Levinson. Measuring the structure of road networks. *Geographical analysis*, 39(3):336–356, 2007.
- [230] Feng Xie and David Levinson. Topological evolution of surface transportation networks. *Computers, Environment and Urban Systems*, 33(3):211–223, 2009.
- [231] Hyejin Youn, Luís M.A. Bettencourt, José Lobo, Deborah Strumsky, Horacio Samaniego, and Geoffrey B. West. The systematic structure and predictability of urban business diversity. *arXiv:1405.3202*, 2014.
- [232] Yacov Zahavi. Traveltime budgets and mobility in urban areas. Technical Report FHWA-PL-8183, Federal Highway Administration, 1974.
- [233] Damián H. Zanette and Susanna C. Manrubia. Role of intermittency in urban development: a model of large-scale city formation. *Physical Review Letters*, 79(3):523, 1997.
- [234] George K. Zipf. *Human behavior and the principle of least effort*. Addison-Wesley Press, 1949.